BJPsych | The British Journal of Psychiatry (2022)
220, 229–245. doi: 10.1192/bjp.2022.16

Using combined environmental–clinical classification models to predict role functioning outcome in clinical high-risk states for psychosis and recent-onset depression

Linda A. Antonucci, Nora Penzel, Rachele Sanfelici, Alessandro Pigoni, Lana Kambeitz-Ilankovic, Dominic Dwyer, Anne Ruef, Mark Sen Dong, Ömer Faruk Öztürk, Katharine Chisholm, Theresa Haidl, Marlene Rosen, Adele Ferro, Giulio Pergola, Ileana Andriola, Giuseppe Blasi, Stephan Ruhrmann, Frauke Schultze-Lutter, Peter Falkai, Joseph Kambeitz, Rebekka Lencer, Udo Dannlowski, Rachel Upthegrove, Raimo K. R. Salokangas, Christos Pantelis, Eva Meisenzahl, Stephen J. Wood, Paolo Brambilla, Stefan Borgwardt, Alessandro Bertolino and Nikolaos Koutsouleris the PRONIA Consortium

Background

Clinical high-risk states for psychosis (CHR) are associated with functional impairments and depressive disorders. A previous PRONIA study predicted social functioning in CHR and recent-onset depression (ROD) based on structural magnetic resonance imaging (sMRI) and clinical data. However, the combination of these domains did not lead to accurate role functioning prediction, calling for the investigation of additional risk dimensions. Role functioning may be more strongly associated with environmental adverse events than social functioning.

Aims

We aimed to predict role functioning in CHR, ROD and transdiagnostically, by adding environmental adverse events-related variables to clinical and sMRI data domains within the PRONIA sample.

Method

Baseline clinical, environmental and sMRI data collected in 92 CHR and 95 ROD samples were trained to predict lower versus higher follow-up role functioning, using support vector classification and mixed k-fold/leave-site-out cross-validation. We built separate predictions for each domain, created multimodal predictions and validated them in independent cohorts (74 CHR, 66 ROD).

Results

Models combining clinical and environmental data predicted role outcome in discovery and replication samples of CHR (balanced accuracies: 65.4% and 67.7%, respectively), ROD (balanced accuracies: 58.9% and 62.5%, respectively), and transdiagnostically (balanced accuracies: 62.4% and 68.2%, respectively). The most reliable environmental features for role outcome prediction were adult environmental adjustment, childhood trauma in CHR and childhood environmental adjustment in ROD.

Conclusions

Findings support the hypothesis that environmental variables inform role outcome prediction, highlight the existence of both transdiagnostic and syndrome-specific predictive environmental adverse events, and emphasise the importance of implementing real-world models by measuring multiple risk dimensions.

Keywords

Machine learning; role functioning; personalised psychiatry; psychosis; PRONIA.

Functioning impairments and risk for psychosis and depression

Loss of functioning is linked to reduced quality of life, and they combine to negatively influence the disease course of many psychiatric conditions, especially psychosis and major depression.[1] Early functional deficits are already present in clinical high-risk states (CHRs).[2,3] More specifically, deficits in role functioning (educational and occupational) within the CHR period are particularly relevant because they frequently develop in CHRs irrespective of transition to psychosis, and lead to problems associated with inability to attend school, unemployment, social impairments and lasting financial consequences.[4,5] Notably, although outcomes in social functioning could recently be promisingly predicted by machine learning models[6] constructed on clinical and structural magnetic resonance imaging (sMRI) baseline data in up to 83% of patients in CHRs,[7] outcomes related to role functioning could not accurately be determined, thus calling for a broader investigation of potential predictors. Evidence shows that social and role functioning may be fundamentally different phenomena, differentially linked to symptoms,[8] neurocognitive deficits[9] and adverse outcomes.[10,11] More specifically, recent views posited that role functioning may be more strongly associated with concurrent environmental factors, compared with social functioning.[12,13] This would be coherent with the notion that environmental adverse events during maturational/developmental periods (e.g. trauma experiences, repeated negative social interactions, maladjustments in developmental goals) are central for psychosis,[14] depression[15] and bipolar disorder[16] pathophysiology. Moreover, such adverse events have been associated with brain structure and function alterations.[17–19]

Notably, impairments in role functioning also concern individuals in early illness stages outside the psychosis risk spectrum, such as depression.[20] This is particularly relevant since the CHR may evolve in different psychiatric disorders.[21] Indeed, 35–68% of patients in CHRs develop or maintain non-psychotic disorders,[20,21] thus calling for research on CHRs to broaden the scope of risk estimation to detect not only psychosis, but also other adverse outcomes. Consistently, patients in CHRs often experience affective symptoms, to the extent that 41% of have a depressive disorder.[21,22] Furthermore, studies showed that psychosis risk is detectable also in

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

affective conditions beyond the traditional 'at-risk' construct.[23,24] These findings highlight that multiple conditions are associated with role impairments since their early stages. Thus, the prediction of such functional outcomes should be targeting both CHRs and affective samples, to obtain more realistic, reliable and potentially transdiagnostic prediction models of future risk for further functional impairments and, ultimately, disability, in a more heterogeneous help-seeking population.[7,25,26]

Employing machine learning to identify reliable functioning predictors

So far, research on early identification of patients who subsequently develop psychosis or other adverse outcomes has produced favourable results, yet further improvement is required. Few studies[27,28] have tested the predictive value of environmental adverse events on functional impairments across concurrent psychiatric conditions, as frequently present in psychosis risk syndromes. In this context, machine learning could harness the interacting effects of different risk factors by building prognostic models using multiple data domains, rather than using only one domain at a time. This form of multimodal learning has been shown to improve prognostic/predictive performance in various fields of medicine, such as affective disorders,[29,30] Alzheimer's disease[31] and stroke.[32] Also, in the CHR field, multimodal risk calculators outperformed unimodal prediction models.[28,33–35] Notably, these multimodal risk calculators also showed generalisability,[7,36] even when applied to outcome prediction.[37] Thus, embracing a multimodal predictive approach could facilitate the identification and characterisation of people at risk for adverse functioning outcomes, which might, in turn, lead to differential managements that are tailored on a patient's individual needs and impairments,[38] irrespective of a later transition to psychosis.[38] Parallel to the development of generalisable predictive models, a deeper investigation into the prognostic power of single data domains and how each domain individually influences the final prediction is also of central importance.[6]

Study aim

The aim of this study was therefore to expand existing role functioning prediction models,[7] which operated on clinical and sMRI data, by adding information regarding environmental factors previously associated with psychosis[39–42] and depression.[43–46] We analysed two populations with overlapping courses of functional impairments (i.e. CHR and recent-onset depression (ROD)), drawn from the database of the Personalized Prognostic Tools for Early Psychosis Management study (PRONIA; https://www.pronia.eu/). We hypothesised that, by adding environmental information to the clinical and sMRI data domains, follow-up role functioning impairments could be more accurately predicted in CHR and ROD samples separately, as well as transdiagnostically. As a first step, we investigated the predictive power of environmental factors occurring before baseline, alone and in combination with clinical (i.e. retrospectively collected scores of social and role functioning) and sMRI data. Then, we investigated the models' transdiagnostic potential and generalisability to unseen individuals. We evaluated the predictive importance of each environmental variable in the respective predictive models, and then investigated whether the best multimodal predictive model generalised to the prediction of other clinically relevant trajectories. As a final step, we employed multivariate regression techniques to assess whether the environmentally determined predictions of role functioning were associated with the clinical and sMRI data domains, and could therefore act *via* clinical vulnerability or sMRI abnormalities.

Method

Sample determination

Individuals were recruited within the European Union's Seventh Framework Programme project PRONIA.[7] The cohort is divided based on the date of recruitment in CHR ($n = 92$) and ROD ($n = 95$) discovery samples (i.e. individuals recruited between February 2014 and May 2017, at seven sites; Table 1), for model generation, and in CHR ($n = 74$) and ROD ($n = 66$) replication samples, for generalisability assessments (i.e. individuals recruited after May 2017 and July 2019 at the same seven discovery sites, plus three new sites; Supplementary Appendix 1 and Supplementary Table 1 available at https://doi.org/10.1192/bjp.2022.16). Individuals meeting the criteria for CHR or ROD were recruited according to internationally established diagnostic criteria;[7] 20% of CHR and 16% of ROD individuals were recruited at the three new replication sites.

For all individuals, baseline sMRI and environmental information, as well as baseline and follow-up social and role functioning scores between the 6- and 12-month timepoints of the study (clinical data), were available. Written informed consent was obtained from all participants. The Global Functioning: Role (GF:R) scale[47] was used to define lower versus higher role functioning at a literature-based threshold,[7] using the participants' latest examination within the 6- to 12-month follow-up period. Based on the literature,[47] a score of >7 points indicated higher outcome and a score of ≤7 points indicated lower outcome, as a score of 7 points marks initial mild, but already persistent, role functioning impairment.

Two-sample *t*-tests, *z*-tests and chi-squared tests were used to investigate potential across-sites demographic and clinical baseline differences in CHR and ROD. Furthermore, we investigated the prevalence comparisons of DSM-IV-TR diagnoses in CHR and ROD with lower versus higher role functioning at baseline (T0) and follow-up examinations 9 months after baseline (T1), through chi-squared tests. *P*-values were group- and timepoint-wise false discovery rate (FDR)-corrected ($\alpha = 0.05$).

Unimodal classifiers

The combined clinical and sMRI role functioning prediction models reported in the previous study from our group,[7] which we aimed to expand by adding environmental information, informed the present analysis with respect to the choice of predictors and the machine learning pipelines for sMRI and clinical classifiers, which were not altered in any part. The environmental classifier was not informed from the previous study,[7] except for the machine learning pipeline, which was implemented consistently with the clinical and sMRI ones. However, it should be noted that our samples do not completely overlap with those used in that study, because 24 CHR and 25 ROD individuals in the discovery samples skipped the environmental assessment (see Supplementary Appendix 1, Section 1). Therefore, only individuals with environmental assessments (besides clinical and sMRI) were retained in the present study. Based on this rationale, we trained the following classifiers:

(a) A 'clinical' classifier based on eight baseline Global Functioning: Social (GF:S) and GF:R scores (i.e. highest social and role functioning lifetime score, highest and lowest social and role functioning scores in the past year, and current social and role functioning scores), based on the Global Functioning Scale.[47]
(b) An 'environmental' classifier, using six summary scores derived from the Childhood Trauma Questionnaire (CTQ[45]), the Bullying Scale for Adults,[48] and time-window scores (childhood, early adolescence, late adolescence and adulthood) of

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

Table 1 Study-associated, sociodemographic, physical, clinical, functional and environmental differences at baseline in discovery individuals with clinical high-risk states, and in individuals with recent-onset depression, with lower versus higher Global Functioning: Role scale outcomes at follow-up

	Follow-up							
	CHR				ROD			
Characteristic	Lower GF:R	Higher GF:R	$t/z/\chi^2$	*P*-value	Lower GF:R	Higher GF:R	$t/z/\chi^2$	*P*-value
Sample sizes and study variables								
Total	52	40			48	47		
Participants per site, *n* (%)								
Munich	16 (30.8)	10 (25)	$\chi^2_6 = 9.38$	0.15	18 (56.3)	14 (27.7)	$\chi^2_6 = 7.32$	0.30
Milan	5 (9.6)	1 (2.5)			2 (4.2)	1 (2.1)		
Basel	5 (9.6)	9 (22.5)			5 (12.5)	5 (8.5)		
Cologne	6 (11.5)	4 (10)			6 (16.7)	11 (23.4)		
Birmingham	4 (7.7)	8 (20)			3 (6.3)	9 (19.1)		
Turku	10 (19.2)	3 (7.5)			6 (12.5)	3 (6.4)		
Udine	6 (11.5)	5 (12.5)			6 (12.5)	6 (12.8)		
Participants examined post-enrolment, number per month								
6	0	0	$\chi^2_1 = 0.06$	0.81	1	0	$\chi^2_2 = 2.13$	0.38
9	34	28			38	33		
12	18	12			9	13		
Sociodemographic data								
Age, mean (s.d.), years	24.1 (5.3)	24 (5.2)	$t_{85} = 01$	0.90	25.2 (5.8)	27.4 (6.1)	$t_{93} = -2.12$	0.04
Male, *n* (%)	27 (51.9)	19 (47.5)	$\chi^2_1 = 0.04$	0.83	24 (50)	16 (34)	$\chi^2_1 = 1.87$	0.17
Edinburgh Handedness Score, mean (s.d.)	59.1 (68.7)	73.6 (46.2)	$t_{82} = -1.16$	0.25	79.9 (34.7)	76.1 (46.8)	$t_{83} = 0.44$	0.66
Education, mean (s.d.), years	13.3 (2.6)	14.1 (3.5)	$t_{68} = -1.20$	0.23	14.6 (2.9)	15.3 (2.9)	$t_{92} = -1.27$	0.21
Educational years repeated, mean (s.d.), years	0.6 (1.0)	0.2 (0.4)	$t_{67} = 2.73$	0.01	0.2 (0.5)	0.7 (1.9)	$t_{51} = -1.66$	0.10
Having a partnership most of the time in the year before study inclusion, *n* (%)	12 (23.1)	7 (17.5)	$\chi^2_1 = 0.11$	0.73	9 (18.8)	16 (34)	$\chi^2_1 = 1.96$	0.16
Population density in living area, mean (s.d.), habitants/km^2	3110.6 (2387.5)	3242.5 (2600.4)	$t_{78} = -0.25$	0.80	3369.6 (2280.2)	3284.1 (2312.1)	$t_{92} = 0.18$	0.86
Number of hospital admissions, mean (s.d.)	0.5 (0.6)	0.4 (0.7)	$t_{80} = 0.78$	0.44	0.7 (0.6)	0.6 (0.6)	$t_{93} = 0.74$	0.46
Clinical high-risk state inclusion criteria								
Schizotypal personality disorder present, *n* (%)	5 (9.6)	1 (2.5)	$\chi^2_1 = 0.89$	0.35	0	0	Not applicable	Not applicable
First-degree relatives with psychosis, *n* (%)	4 (7.7)	10 (25)	$\chi^2_1 = 3.99$	0.05	1 (2.1)	0	$\chi^2_1 = 0$	<0.99
30% Loss of global functioning compared with highest levels in the year before study inclusion, *n* (%)	17 (32.7)	8 (20)	$\chi^2_1 = 1.25$	0.26	9 (18.8)	5 (10.6)	$\chi^2_1 = 0.68$	0.41
Genetic risk disability schizotypal personality disorder criterion criteria met, *n* (%)	5 (9.6)	8 (20)	$\chi^2_1 = 1.24$	0.26	0	0	Not applicable	Not applicable
Cognitive disturbances criteria met, *n* (%)	27 (51.9)	23 (57.5)	$\chi^2_1 = 0.10$	0.75	0	0	Not applicable	Not applicable
Attenuated psychotic symptoms criteria met, *n* (%)	38 (73.1)	21 (52.5)	$\chi^2_1 = 3.32$	0.07	0	0	Not applicable	Not applicable
Brief limited intermittent psychotic symptoms criteria met, *n* (%)	1 (1.9)	3 (7.5)	$\chi^2_1 = 0.62$	0.43	0	0	Not applicable	Not applicable
Global Assessment of Functioning score at baseline, mean (s.d.)								
Disability, highest lifetime score	78.4 (9.1)	82 (5.6)	$t_{86} = -2.32$	0.02	80.3 (9.3)	82.9 (7.3)	$t_{89} = -1.54$	0.13
Symptoms, highest lifetime score	78.6 (8.9)	81.1 (8.1)	$t_{87} = -1.40$	0.16	81.8 (7.6)	83.2 (7.4)	$t_{93} = -0.90$	0.37
Disability, score in past year	65.0 (13.2)	71.5 (10.1)	$t_{90} = -2.65$	0.01	68.8 (14)	75 (12.7)	$t_{92} = -2.26$	0.03

(*Continued*)

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

Table 1 (*Continued*)

	Follow-up							
	CHR				ROD			
Characteristic	Lower GF:R	Higher GF:R	$t/z/\chi^2$	*P*-value	Lower GF:R	Higher GF:R	$t/z/\chi^2$	*P*-value
Symptoms, score in past year	63.3 (12.6)	70 (9.8)	$t_{90} = -2.83$	0.01	71.5 (10.6)	73.9 (12.0)	$t_{91} = -1.06$	0.29
Disability, score in past month	53 (12.7)	62 (15)	$t_{76} = -3.07$	0.002	52.6 (13.7)	61.4 (15.4)	$t_{91} = -2.94$	0.004
Symptoms, score in past month	53.9 (10.7)	57.5 (10.8)	$t_{84} = -1.59$	0.12	55.1 (12.1)	58.4 (13.0)	$t_{92} = -1.25$	0.21
Global Functioning: Social scale, mean (s.d.) score								
Highest lifetime score	7.6 (0.9)	8.2 (0.6)	$z = -3.34$	0.001	8 (0.9)	8.3 (0.8)	$z = -2.01$	0.05
Highest score in past year	6.8 (1.4)	7.7 (0.7)	$z = -3.53$	<0.001	7.3 (1.0)	7.5 (1.1)	$z = -1.11$	0.27
Baseline score	6.2 (1.4)	7.0 (1.1)	$z = 3.20$	0.002	6.3 (1.2)	6.9 (1.3)	$z = -1.98$	0.05
Global Functioning: Role scale, mean (s.d.) score								
Highest lifetime score	7.7 (0.9)	8.3 (0.7)	$z = -3.74$	<0.001	8.2 (0.8)	8.5 (0.9)	$z = -1.85$	0.07
Highest score in past year	6.8 (1.2)	7.8 (0.9)	$z = -4.75$	<0.001	7.4 (1.1)	8 (1.2)	$z = -2.41$	0.02
Baseline score	5.7 (1.2)	7.0 (1.3)	$z = -5.07$	<0.001	6.1 (1.5)	6.8 (1.6)	$z = -2.26$	0.03
Standardized Interview for Prodromal Symptoms score at baseline, mean (s.d.) scores								
Unusual thought content or delusional ideas	2.6 (1.5)	2.5 (1.5)	$z = 0.28$	0.78	1 (1.1)	0.9 (1.0)	$z = 0.40$	0.69
Suspiciousness or persecutory ideas	2.0 (1.8)	1.9 (2.1)	$z = 0.35$	0.73	0.2 (0.7)	0.2 (0.5)	$z = 0.29$	0.77
Grandiosity	0.3 (1)	0.3 (0.9)	$z = 0.01$	0.99	0.0 (0.1)	0.0 (0.3)	$z = -0.46$	0.65
Perceptual abnormalities	2.1 (1.9)	1.8 (1.5)	$z = 0.89$	0.38	0.7 (1)	0.8 (1.3)	$z = -0.52$	0.60
Disorganised communication	1.1 (1.5)	0.6 (1.1)	$z = 1.82$	0.07	0.3 (0.6)	0.0 (0.2)	$z = 2.65$	0.01
Social anhedonia	2.1 (1.8)	1.5 (1.8)	$z = 1.56$	0.12	2 (1.8)	1.2 (1.6)	$z = 2.06$	0.04
Avolition	2.5 (1.5)	1.9 (1.8)	$z = 1.67$	0.10	2.7 (1.6)	2.2 (1.8)	$z = 1.35$	0.18
Expression of emotion	1.2 (1.6)	0.8 (1.3)	$z = 1.50$	0.14	1 (1.2)	0.5 (1)	$z = 2.09$	0.04
Experience of emotions and self	1.4 (1.5)	1.2 (1.7)	$z = 0.66$	0.51	1.3 (1.4)	1.1 (1.6)	$z = 0.61$	0.54
Ideational richness	0.6 (1.3)	0.1 (0.5)	$z = 2.45$	0.02	0.3 (0.8)	0.0 (0.1)	$z = 1.99$	0.05
Occupational functioning	3.0 (1.8)	1.4 (1.7)	$z = 4.45$	<0.001	2.7 (1.7)	1.9 (1.7)	$z = 2.37$	0.02
Beck Depression Inventory sum score	26.2 (12.7)	22.3 (11.3)	$t_{85} = 1.54$	0.13	23.5 (13.7)	26.3 (15.3)	$t_{91} = -0.94$	0.35
Positive and Negative Symptoms Scale, mean (s.d.) scores								
Total	53.3 (15.5)	45.4 (9.4)	$t_{84} = 3.01$	0.003	48 (8.9)	45 (9.6)	$t_{92} = 1.59$	0.11
Positive sum	10.4 (3.1)	9.5 (2.8)	$t_{87} = 1.35$	0.18	7.7 (1.3)	7.6 (1.0)	$t_{88} = 0.36$	0.72
Negative sum	14.6 (6.6)	9.8 (3)	$t_{72} = 4.61$	<0.001	13.0 (4.6)	11.3 (4.2)	$t_{91} = 1.96$	0.05
General sum	28.9 (7.8)	26 (6.1)	$t_{88} = 1.0$	0.05	27.3 (5.7)	26.1 (6)	$t_{92} = 0.97$	0.34
Childhood Trauma Questionnaire, mean (s.d.) scores								
Raw scores	32.94 (4.96)	34 (6.32)	$z = -0.83$	0.41	33.48 (7.57)	34.12 (7.08)	$z = -0.83$	0.41
Deviation scores	0.08 (0.48)	0.26 (0.57)	$z = -1.52$	0.13	0.16 (0.71)	0.22 (0.66)	$z = -1.52$	0.13
Premorbid Adjustment Scale, mean (s.d.) scores								
Childhood								
Raw score (total)	0.16 (0.12)	0.2 (0.15)	$z = -1.12$	0.27	0.31 (0.18)	0.27 (0.2)	$z = -0.40$	0.69
Deviation score (total)	0.36 (1.2)	0.42 (1.44)	$z = -0.20$	0.84	0.64 (1.42)	0.35 (1.32)	$z = -0.40$	0.69
Early adolescence								
Raw score (total)	0.26 (0.18)	0.27 (0.17)	$z = -0.46$	0.65	0.17 (0.12)	0.19 (0.15)	$z = 1.11$	0.27
Deviation score (total)	1.23 (2)	1.19 (1.84)	$z = 0.11$	0.91	1.78 (1.98)	1.3 (2.19)	$z = 0.1$	0.32
Late adolescence								
Raw score (total)	0.26 (0.18)	0.23 (0.14)	$z = 0.74$	0.46	0.29 (0.19)	0.26 (0.17)	$z = 0.67$	0.51
Deviation score (total)	0.99 (1.6)	0.8 (1.36)	$z = 0.61$	0.54	1.31 (1.7)	1.08 (1.54)	$z = 1.11$	0.27

(*Continued*)

Table 1 *(Continued)*

Characteristic	Follow-up							
	CHR				ROD			
	Lower GF:R	Higher GF:R	$t/z/\chi^2$	*P*-value	Lower GF:R	Higher GF:R	$t/z/\chi^2$	*P*-value
Adulthood								
Raw score (total)	0.2 (0.15)	0.17 (0.11)	$z = 0.99$	0.33	0.19 (0.11)	0.13 (0.1)	$z = 2.44$	0.02
Deviation score (total)	1.39 (1.69)	0.24 (0.99)	$z = 3.82$	<0.001	0.86 (1.26)	0.23 (1.11)	$z = 0.67$	0.51
Bullying Scale, mean (s.d.) scores								
Raw score (total)	35.56 (25.45)	31.2 (24.64)	$z = 0.83$	0.41	36.94 (27.85)	35.68 (24.24)	$z = 0.23$	0.82
Deviation score (total)	2.88 (3.06)	2.99 (2.71)	$z = -0.18$	0.86	3.31 (3.22)	3.16 (2.80)	$z = 0.23$	0.82
World Health Organization Quality of Life – Brief Questionnaire, mean (s.d.) scores								
Physical health	18.56 (2.99)	19.55 (3.16)	$t_{77} = -1.48$	0.14	21.14 (3.85)	20.89 (3.83)	$t_{87} = 0.31$	0.76
Psychological health	16.19 (3.05)	17.61 (3.36)	$t_{75} = -2.02$	0.04	16.56 (2.98)	17.17 (3.34)	$t_{87} = -0.92$	0.36
Social relationships	8.94 (2.87)	9.47 (2.36)	$t_{84} = -0.9$	0.34	8.04 (3.67)	9.13 (2.79)	$t_{88} = -1.63$	0.11
Environment	28.23 (4.09)	29.24 (4.36)	$t_{77} = -1.09$	0.28	25.75 (10.06)	28.57 (6.34)	$t_{90} = -1.64$	0.10

Significance was defined at $\alpha = 0.05$. CHR, clinical high-risk; ROD, recent-onset depression; GF:R, Global Functioning: Role scale.

the Premorbid Adjustment Scale (PAS[49]), which measures the relative level of harmony between an individual's needs and environmental characteristics and requests.[50] Notably, although the PAS does not strictly measure adverse events, its derived scores reflect how environmental challenges and risk factors may modulate the capacity of people to adjust in different periods of life.[49] All summary scores entering the algorithm were derived by normalising total raw scores by using the published psychometric norms of each instrument (Supplementary Appendix 1, Section 2).

(c) An 'sMRI' classifier, including baseline whole-brain grey matter volume (GMV) individual data. We employed open-source CAT12 toolbox (version r1155 for Linux; Christian Gaser, University of Jena, Germany; see http://dbm.neuro.uni-jena.de/cat12/) to pre-process and analyse individual GMV maps. Detailed consortium-wise pre-processing and site correction procedures of MRI data is reported in Supplementary Appendix 1, Section 3 and elsewhere.[7]

Classifiers were based either on continuous variables (i.e. environment and sMRI) or ordinal variables (i.e. clinical), which were treated as continuous variables based on previous literature.[51] All predictor assessments were made without knowledge of outcome data, and the outcome label was determined without knowledge of predictor information. Two-sample *t*-tests and *z*-tests were used to investigate potential clinical and environmental differences between CHR and ROD with regards to lower versus higher role functioning ($P < 0.05$). Results for discovery cohorts are reported in Table 1, and results for the replication cohort are reported in Supplementary Table 1. For MRI, we ran checks to rule out any role functioning or site effects, as well as their interaction, on GMV estimates. The results of these checks highlighted the absence of any main effect of role functioning, site and their interaction, on GMV maps (all $P > 0.05$, family-wise error-corrected k = 10; Supplementary Appendix 1, Section 4).

Machine learning pipeline

The overall analytic strategy (Supplementary Fig. 1) was to first quantify the unimodal prognostic performance of each classifier (clinical, environmental, sMRI), and then to understand whether environmental information would improve prediction performance when combined with the clinical model and/or the sMRI model. Therefore, for each cohort (CHR and ROD), we built six machine learning models to predict higher versus lower GF:R outcome: three using unimodal classifiers and three using combinations of individual classifiers (multimodal classifiers). To facilitate comparability and interpretation of our findings, both for unimodal and multimodal classifiers, we chose to implement the same machine learning pipelines reported to generate the combined clinical and sMRI prediction models we aimed at expanding.[7] With this aim, we implemented a mixed inner k-fold/outer leave-site-out[7] cross-validation strategy based on our machine learning platform NeuroMiner, version 1.0 for Linux (Nikolaos Koutsouleris, Munich, Germany; see https://github.com/neurominer-git). Per each population (CHR, ROD), we obtained unimodal risk calculator predictions based on environmental, clinical and sMRI baseline features. On the basis of these unimodal predictions, we built the three multimodal classifiers described above, using stacked generalisation (Supplementary Appendix 1, Section 5).[52] We purposefully did not investigate the joint predictive ability of clinical and sMRI classifiers, as this was already explicitly addressed in a recent publication from our group.[7] To measure the discriminative utility of the input variables within each unimodal classifier, we computed the probability of being selected for classification purposes within the inner cross-validation loop for each feature,[27] following a

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

Table 2 Performance of all unimodal and multimodal classifiers tested in clinical high-risk, recent-onset depression and the pooled sample, for global functioning role outcome at follow-up

	True positives	False positives	True negatives	False negative	Accuracy	Specificity	Sensitivity	FPR	PPV	NPV	AUC (95% CI)	BAC (minimum–maximum) [s.d.]	Permutation-based significance
CHR: global functioning role prediction													
Environmental: discovery CHR	34	13	27	18	66.3	67.5	65.4	32.5	72.3	60.0	0.63 (0.52–0.74)	66.4 (40.0–81.7) [13.0]	**0.01**
Environmental: replication CHR, all sites	60	14	0	0	81.0	0	100	100	81.0	Not assessed	0.51 (0.39–0.63)	50.0 (50.0–50.0) [0.0]	Not assessed
Clinical: discovery CHR	31	10	30	21	66.3	75.0	59.6	25.0	75.6	58.8	0.76 (0.66–0.86)	67.3 (50.0–90.0) [16.0]	**0.04**
Clinical: replication CHR, all sites	38	4	10	22	64.8	71.4	63.3	28.5	90.4	31.2	0.76 (0.66–0.86)	67.3 (61.9–74.8) [3.3]	Not assessed
sMRI: discovery CHR	28	17	23	24	48.9	53.8	42.5	57.5	54.9	41.5	0.51 (0.39–0.63)	48.2 (30.0–80.0) [12.6]	0.63
sMRI: replication CHR, all sites	26	12	14	60	18.9	100	0	0	Not assessed	18.9	0.30 (0.19–0.41)	50.0 (42.4–62.1) [3.3]	Not assessed
Three modalities: environmental + clinical + sMRI: discovery CHR	35	30	10	17	70.7	67.3	75.0	25.0	77.8	63.8	0.75 (0.65–0.85)	71.2 (51.3–90.0) [12.1]	**0.01**
Three modalities: environmental + clinical + sMRI: replication CHR, all sites	49	9	5	11	72.9	35.9	81.6	64.3	84.4	31.2	0.70 (0.59–0.81)	58.7 (48.2–69.0) [4.6]	Not assessed
Two modalities: environmental + clinical: discovery CHR	29	30	10	23	64.1	55.8	75.0	25.0	74.4	56.6	0.76 (0.66–0.86)	65.4 (50.0–100.0) [15.6]	**0.04**
Two modalities: environmental + clinical: replication CHR, all sites	47	6	8	13	74.3	57.1	78.3	42.8	88.6	38.1	0.76 (0.66–0.86)	67.7 (52.7–71.5) [5.2]	Not assessed
Two modalities: environmental + sMRI: discovery CHR	31	23	17	21	58.7	59.6	57.5	42.5	64.6	52.3	0.57 (0.45–0.69)	58.6 (20.0–81.7) [17.4]	0.11
Two modalities: environmental + sMRI: replication CHR, all sites	59	14	0	1	79.7	0	98.3	100	80.8	0	0.30 (0.19–0.41)	49.2 (37.7–54.0) [3.5]	Not assessed
ROD: global functioning role prediction													
Environmental: discovery ROD	24	17	30	24	56.8	63.8	50.0	36.2	58.5	55.6	0.53 (0.41–0.65)	56.9 (40.2–75.0) [11.4]	0.11
Environmental: replication ROD, all sites	9	2	20	35	43.9	90.9	20.4	9.1	81.8	36.3	0.56 (0.44–0.68)	55.6 (47.7–56.6) [1.9]	Not assessed
Clinical: discovery ROD	30	22	25	18	57.9	53.2	62.5	46.8	57.7	58.1	0.62 (0.51–0.73)	57.8 (41.7–79.2) [11.8]	0.12
Clinical: replication ROD, all sites	38	10	12	6	75.5	54.5	86.3	45.5	79.1	66.6	0.81 (0.72–0.90)	70.5 (63.6–79.5) [5.3]	Not assessed
sMRI: discovery ROD	24	26	21	24	47.4	44.7	50.0	55.3	48.0	46.7	0.49 (0.37–0.61)	47.3 (16.7–75.0) [17.2]	0.72
sMRI: replication ROD, all sites	40	19	3	4	65.2	13.6	90.9	86.3	67.7	42.8	0.53 (0.41–0.65)	52.2 (44.3–56.8) [3.4]	Not assessed
Three modalities: environmental + clinical + sMRI: discovery ROD	29	23	24	19	55.8	51.1	60.4	48.9	55.8	55.8	0.54 (0.42–0.66)	55.7 (33.3–67.6) [10.1]	0.16
Three modalities: environmental + clinical + sMRI: replication ROD, all sites	22	6	16	22	57.6	72.7	50.0	27.2	78.5	42.1	0.53 (0.41–0.65)	61.3 (44.3–56.8) [6.3]	Not assessed
Two modalities: environmental + clinical: discovery ROD	30	21	26	18	58.9	55.3	62.5	44.7	58.8	59.1	0.60 (0.49–0.71)	58.9 (40.2–75.0) [11.4]	**0.04**
Two modalities: environmental + clinical: replication ROD, all sites	21	17	5	23	57.5	77.2	47.7	22.7	80.7	42.5	0.72 (0.62–0.82)	62.5 (47.7–78.4) [8.3]	Not assessed
Two modalities: environmental + sMRI: discovery ROD	25	22	25	23	52.6	53.2	52.1	46.8	53.2	52.1	0.52 (0.40–0.64)	52.6 (25.0–67.6) [11.5]	0.31

(*Continued*)

Table 2 (*Continued*)

	True positives	False positives	True negatives	False negative	Accuracy	Specificity	Sensitivity	FPR	PPV	NPV	AUC (95% CI)	BAC (minimum–maximum) [s.d.]	Permutation-based significance
Two modalities: environmental + sMRI: replication ROD, all sites	17	5	17	27	51.5	77.2	38.6	2.7	77.2	38.6	0.53 (0.41–0.65)	57.9 (44.3–62.5) [5.0]	Not assessed
Transdiagnostic (CHR + ROD): global functioning role prediction													
Environmental: discovery (CHR + ROD)	45	26	61	55	56.6	70.1	45.0	29.8	63.3	52.5	0.58 (0.50–0.66)	57.7 (40.0–78.6) [9.2]	0.06
Environmental: replication (CHR + ROD), all sites	16	4	32	88	34.3	88.8	15.4	11.1	80.0	26.6	0.58 (0.50–0.66)	52.1 (48.3–60.5) [3.1]	Not assessed
Environmental: replication (only CHR), all sites	11	2	12	49	31.1	85.7	18.3	4.2	84.6	19.6	0.72 (0.65–0.79)	52.0 (45.4–55.8) [2.4]	Not assessed
Environmental: replication (only ROD), all sites	5	2	20	39	57.8	90.9	11.3	9.1	71.4	33.8	0.68 (0.60–0.76)	51.1 (47.7–55.7) [1.9]	Not assessed
Clinical: discovery (CHR + ROD)	61	27	60	39	64.7	68.9	61.0	31.0	69.3	60.6	0.72 (0.65–0.79)	64.9 (45.4–87.7) [13.7]	**0.02**
Clinical: replication (CHR + ROD), all sites	67	9	27	37	67.1	75.0	64.4	25.0	88.1	42.1	0.74 (0.67–0.81)	69.7 (63.2–74.5) [3.5]	Not assessed
Clinical: replication (only CHR), all sites	39	6	8	21	63.5	57.1	65.0	2.8	86.6	27.6	0.75 (0.68–0.82)	61.1 (59.4–73.2) [5.2]	Not assessed
Clinical: replication (only ROD), all sites	28	3	19	16	71.2	86.3	63.6	13.6	90.3	54.3	0.72 (0.65–0.79)	75.0 (58.0–76.1) [4.2]	Not assessed
sMRI: discovery (CHR + ROD)	42	35	49	58	41.8	41.6	42.0	58.3	46.1	37.6	0.40 (0.32–0.48)	41.8 (24.5–85.7) [13.3]	0.5
sMRI: replication (CHR + ROD), all sites	56	15	21	48	55.0	58.3	53.8	41.6	78.8	30.4	0.59 (0.51–0.67)	56.0 (46.0–61.6) [3.8]	Not assessed
sMRI: replication (only CHR), all sites	31	6	8	29	52.7	57.1	51.6	42.8	83.7	21.6	0.64 (0.56–0.72)	54.4 (33.7–73.9) [10.0]	Not assessed
sMRI: replication (only ROD), all sites	25	9	13	19	57.6	59.1	56.8	40.9	73.6	40.6	0.55 (0.47–0.63)	57.9 (44.3–60.2) [3.6]	Not assessed
Three modalities: environmental + clinical + sMRI: discovery (CHR + ROD)	58	31	56	42	60.9	64.3	58.0	35.6	65.1	57.1	0.67 (0.59–0.75)	61.1 (45.4–85.7) [10.5]	**0.04**
Three modalities: environmental + clinical + sMRI: replication (CHR + ROD), all sites	61	7	29	43	64.2	80.5	58.6	19.4	89.7	40.2	0.75 (0.68–0.82)	69.6 (64.9–73.5) [2.1]	Not assessed
Three modalities: environmental + clinical + sMRI: replication (only CHR), all sites	38	4	10	22	64.4	71.4	63.3	28.5	90.4	31.2	0.77 (0.70–0.84)	67.4 (60.2–78.9) [4.6]	Not assessed
Three modalities: environmental + clinical + sMRI: replication (only ROD), all sites	23	3	19	21	63.3	86.3	2.2	13.6	88.4	47.5	0.72 (0.65–0.79)	69.3 (63.6–75.0) [2.7]	Not assessed
Two modalities: environmental + clinical: discovery (CHR + ROD)	57	59	28	43	62.0	67.8	57.0	32.1	67.0	57.8	0.71 (0.64–0.78)	62.4 (45.4–85.7) [12.8]	**0.03**
Two modalities: environmental + clinical: replication (CHR + ROD), all sites	61	8	28	43	63.5	77.7	58.6	22.2	88.4	39.4	0.74 (0.7–0.8)	68.2 (65.7–72.2) [2.0]	Not assessed

(*Continued*)

Table 2 *(Continued)*

	True positives	False positives	True negatives	False negative	Accuracy	Specificity	Sensitivity	FPR	PPV	NPV	AUC (95% CI)	BAC (minimum–maximum) [s.d.]	Permutation-based significance
Two modalities: environmental + clinical: replication (only CHR), all sites	36	5	9	24	60.8	64.2	60.0	35.7	87.8	27.3	0.76 (0.67–0.81)	62.1 (57.3–72.0) [4.6]	Not assessed
Two modalities: environmental + clinical: replication (only ROD), all sites	25	3	19	19	66.6	86.3	56.8	13.6	89.2	50.0	0.73 (0.66–0.80)	71.5 (64.8–77.3) [3.3]	Not assessed
Two modalities: environmental + sMRI: discovery (CHR + ROD)	39	32	55	61	50.2	63.2	39.0	36.7	54.9	47.4	0.49 (0.41–0.57)	51.1 (41.3–85.7) [11.7]	0.4
Two modalities: environmental + sMRI: replication (CHR + ROD), all sites	26	3	33	78	42.1	91.6	25.0	8.3	89.6	29.7	0.62 (0.54–0.70)	58.3 (48.0–58.9) [2.6]	Not assessed
Two modalities: environmental + sMRI: replication (only CHR), all sites	15	2	12	45	36.4	85.7	25.0	14.3	88.2	21.1	0.58 (0.50–0.66)	55.3 (47.5–62.9) [3.5]	Not assessed
Two modalities: environmental + sMRI: replication (only ROD), all sites	11	1	21	33	48.4	95.4	25.0	4.5	91.6	38.8	0.58 (0.50–0.66)	60.2 (46.6–63.6) [3.9]	Not assessed

Results are reported for both discovery and replication cohorts. Model significances were assessed by computing BAC in 1000 random label permutations and comparing them with the observed BAC of the respective model. *P*-values were adjusted for multiple comparisons, using the false discovery rate (FDR). FDR-corrected significant models are reported in the significance column in bold. Significances could not be calculated for any replication analysis, as the out-of-sample validation mode of NeuroMiner, different from the discovery mode, does not allow us to calculate significance for models that are generated in one cohort and then applied to another sample. For these analyses, 'not assessed' is reported in the table. The range from minimum to maximum BAC across all Outer Cross-Validation (CV2) folds. FPR, false positive rate; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; BAC, balanced accuracy; CHR, clinical high-risk; sMRI, structural magnetic resonance imaging; ROD, recent-onset depression.

forward feature selection procedure that occurred only in the inner cycle (CV1) of training data. Of note, this procedure was performed only for significant unimodal classifiers (see Results). A detailed description of our machine learning pipeline is provided in Supplementary Appendix 1, Sections 5 and 6. *P*-values reflecting the permuted significance of models (Supplementary Appendix 1, Section 6) are reported in Table 2. Permutation-based pairwise comparisons between discovery unimodal and multimodal classifier performance are reported in Supplementary Table 5. As a check, we have calculated the expected calibration error (ECE)[25,53] to estimate calibration for the models achieving best accuracy and generalisability in our discovery cohorts (see Results). The ECE was relatively low, although not perfect (mean ECE = 0.21). Methods and results of this check are fully described in Supplementary Appendix 1, Section 10 and Supplementary Fig. 2.

Assessment of generalisability

Transdiagnostic potential of risk calculators

To investigate whether the hypothesised unimodal and multimodal risk calculators that we computed separately for CHR and ROD could have transdiagnostic potential, we repeated the same pre-processing, training and testing pipeline (Supplementary Appendix 1, Sections 5 and 6) on the combined CHR-ROD population, comprising 187 individuals (92 CHR, 95 ROD).

Validation of risk calculators

To test for the generalisability of all prognostic models derived from CHR, ROD and the pooled CHR and ROD sample, we validated CHR discovery models in the CHR replication cohort ($n = 74$); ROD discovery models in the ROD replication cohort ($n = 66$); and transdiagnostic (CHR + ROD) discovery models in the pooled CHR and ROD replication sample ($n = 140$), without any re-training (Supplementary Appendix 1, Section 5).

Despite the employment of leave-site-out cross-validation, we used sanity checks to rule out whether the discovery and validation performance of our unimodal and multimodal classifiers could be affected by any latent site effects. Results of these checks are reported in Supplementary Appendix 1, Section 5.

Prognostic generalisation of risk calculators to clinical trajectories

To assess our role functioning predictor's generalisability to the development of other clinical outcomes over time, we used linear mixed effects models (see Results). We therefore generated trajectories based on three longitudinal timepoints for three clinical readouts (Supplementary Appendix 1, Section 7): number of psychiatric hospital admissions across timepoints; prodromal positive and negative symptoms, drawn from the Structured Interview for Psychosis-Risk Syndromes;[54] and quality of life, drawn from the World Health Organization Quality of Life – Brief Questionnaire.[55] For each clinical variable of interest, baseline, T1 (6–12 months after baseline) and T2 assessment (18 months after baseline) evaluations were entered into the analyses (Supplementary Appendix 1, Section 7). A multiple comparisons correction was carried out with FDR ($\alpha = 0.05$).

Environmental feature knock-out analysis

To quantify the predictive contribution of each of the environmental variables included in the algorithm, we ran new GF:R outcome prediction models based on environmental features, but we removed each of the six features originally included in the individual classifier, one at a time, without altering the original machine learning pipeline employed for the environmental classifier. This led to six independent

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

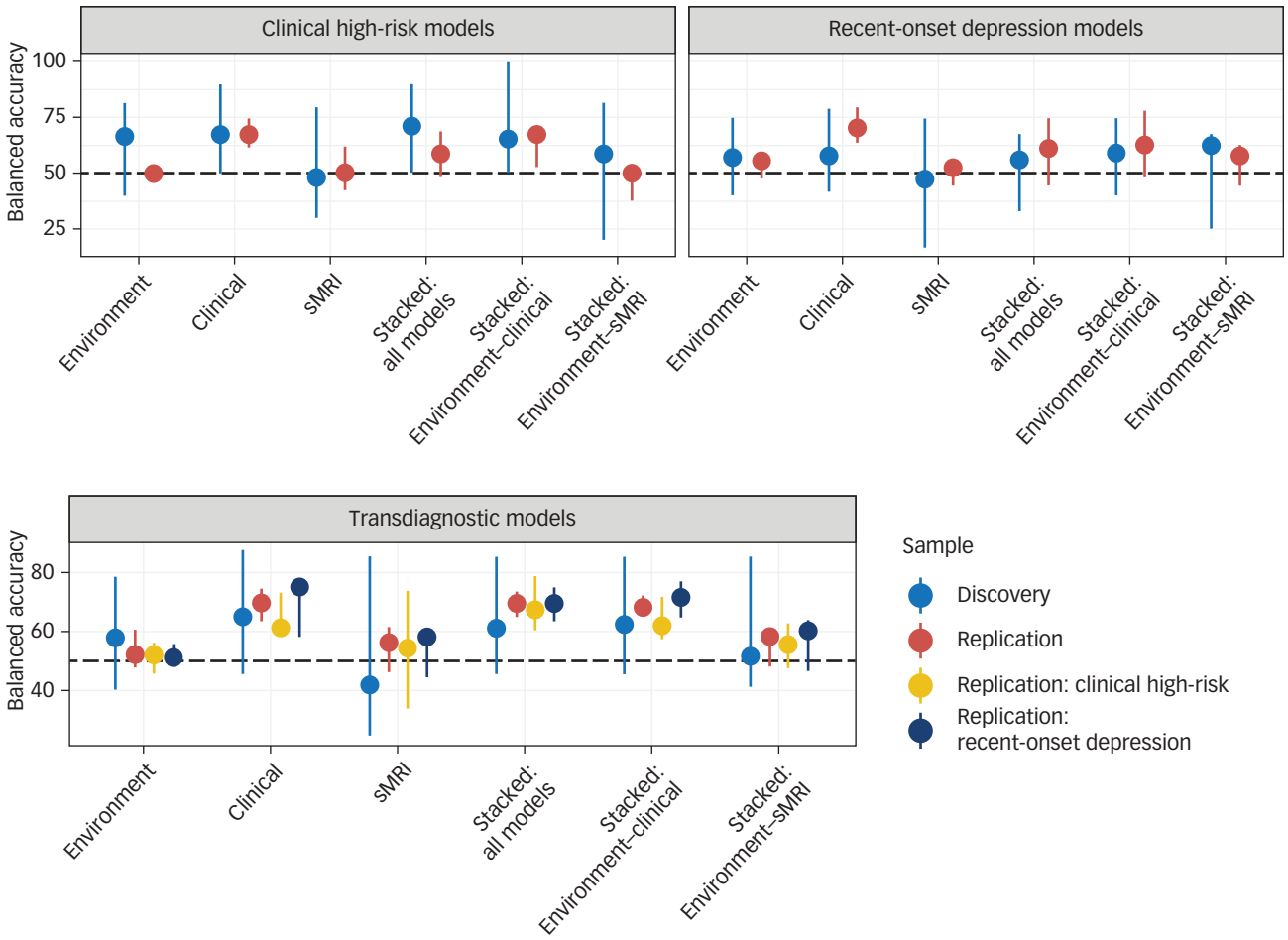

Fig. 1 Plots representing the balanced accuracies (BACs) across models and samples. Dots indicate the mean BAC per model, and bars indicate the range (minimum to maximum) of the BACs per models across all of the cross-validation outer cycle folds. sMRI, structural MRI.

Support Vector Machine analyses for CHR and ROD, each comprising five features.

Investigation of between-classifiers relationships

To understand whether environmentally determined predictions could act via clinical vulnerability or sMRI abnormalities in increasing the risk for worse outcome in CHR and ROD – that is, to preliminarily investigate whether the predictive power of our environmental model on follow-up occupational functioning might be partially explained either by baseline global functioning impairments or baseline sMRI anomalies – we ran support vector regression analyses (Supplementary Appendix 1, Section 9).

Ethic statement

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human patients were approved by the German Clinical Trials Register (DRKS00005042) and approved by the local research ethics committees in each centre.

Results

Demographic, clinical and environmental site-level characteristics are reported separately for CHR and ROD discovery patients in Table 1, and for CHR and ROD replication patients in Supplementary Table 1. Table 3 and Supplementary Table 2 report prevalence comparisons of DSM-IV-TR diagnoses in CHR and ROD samples with lower versus higher GF:R at baseline (T0) and T1 follow-up examinations.

Discovery results

CHR cohort

In the CHR group (Table 2, Fig. 1 and Supplementary Appendix 1, Section 4), only environmental and clinical risk calculators predicted GF:R outcomes significantly better than chance, according to leave-site-out cross-validated balanced accuracy (BAC_{LSOCV}) (environmental model: $BAC_{LSOCV} = 66.4\%$, $P_{FDR} = 0.01$; clinical model: $BAC_{LSOCV} = 67.3\%$, $P_{FDR} = 0.04$) and area under the curve (AUC) (environmental model: 0.63; clinical model: 0.76). The multimodal classifier integrating environmental, clinical and sMRI predictions was more accurate than all unimodal classifiers ($BAC_{LSOCV} = 71.2\%$, $P_{FDR} = 0.01$, AUC = 0.75), followed by the model integrating environmental and clinical predictions ($BAC_{LSOCV} = 65.4\%$, $P_{FDR} = 0.04$, AUC = 0.70). Permutation-based pairwise comparisons between discovery unimodal and multimodal classifier performances are reported in Supplementary Table 5. In the environmental domain, higher deviation scores for premorbid adjustment in adulthood, self-reported bullying victimisation and self-reported experiences of childhood trauma were predictive of poor role functioning outcomes (Fig. 2a). In the clinical

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

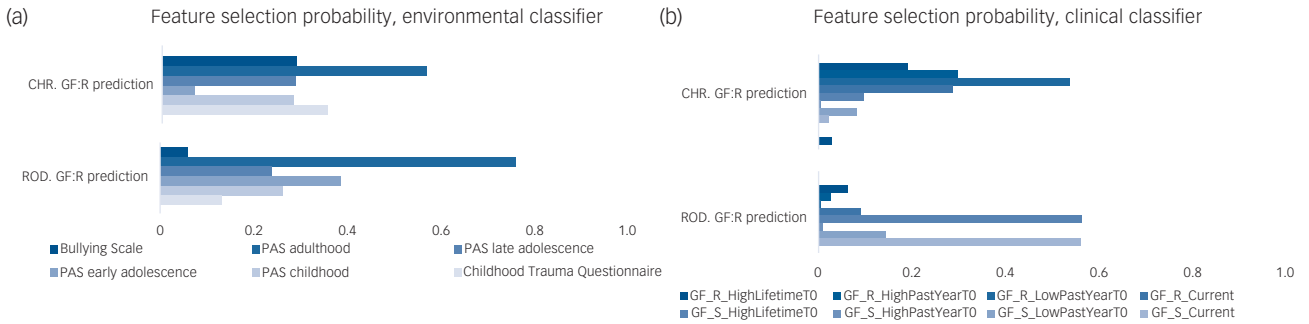

Fig. 2 Probability of each feature for being selected in our mixed k-fold/leave-site-out cross-validation framework by (a) the environmental classifier, (b) the clinical classifier and (c) the sMRI classifier. For (a) and (b), a value of 1 indicates that all models had retained the given variable (Supplementary Appendix 1, Section 5). Feature permutation testing results are reported in Supplementary Table 10. CHR, clinical high-risk; GF:R, Global Functioning: Role; GF:S, Global Functioning: Social; PAS, Premorbid Adjustment Scale; ROD, recent-onset depression; sMRI, structural magnetic resonance imaging; GF_R_HighLifetimeT0, GF:R highest lifetime score measured at T0; GF_R_HighPastYearT0, GF:R highest score in the last year before T0; GF_R_LowPastYearT0, GF:R lowest score in the last year before T0; GF_R_Current, GF:R score at T0 examination; GF_S_HighLifetimeT0 , GF:S highest lifetime score measured at T0; GF_S_HighPastYearT0, GF:S highest score in the last year before T0; GF_S_LowPastYearT0, GF:S lowest score in the last year before T0; GF_S_Current, GF:S score at T0 examination.

domain, lower GF:R lifetime scores predicted poor role functioning outcomes (Fig. 2b).

ROD cohort

In the ROD group (Table 2, Fig. 1 and Supplementary Appendix 1, Section 4), no unimodal risk calculator predicted a GF:R outcome above chance. Only the stacked model integrating environmental and clinical predictions provided significant prediction performance (BAC_{LSOCV} = 58.9%, P_{FDR} = 0.04, AUC = 0.60). Lower GF:R outcomes were predicted by the PAS adulthood, PAS early adolescence and PAS childhood features of the environmental risk calculator (Fig. 2a), as well as by GF:R highest scores during the past year and GF:R highest lifetime scores in the clinical risk calculator (Fig. 2b). Permutation-based pairwise comparisons between discovery unimodal and multimodal classifier performances are reported in Supplementary Table 5.

Assessment of generalisability

Transdiagnostic potential of risk calculators

In the pooled CHR and ROD discovery sample (Table 2 and Fig. 1), the clinical risk calculator predicted a GF:R outcome above chance and with significance (BAC_{LSOCV} = 64.8%, P_{FDR} = 0.02, AUC = 0.72). In the environmental domain, higher deviation scores for PAS adulthood and PAS late adolescence were predictive of poor role functioning outcomes. In the clinical domain, lowest current GF:R score and lowest GF:S score during the past year predicted poor role functioning outcome. The multimodal classifier combining clinical and environmental predictions, as well as the model integrating all unimodal models, performed significantly above chance (combined clinical + environmental model: BAC_{LSOCV} = 62.4%, P_{FDR} = 0.03, AUC = 0.71; combined clinical + enviornmental + sMRI model: BAC_{LSOCV} = 61.1%, P_{FDR} = 0.04, AUC = 0.67).

Validation of risk calculators

The study-group-specific clinical models, as well as multimodal risk calculators combining environmental and clinical data, performed above chance when applied to the respective CHR and ROD replication samples (CHR: BAC_{LSOCV} = 67.3% and AUC = 0.76 for the clinical model, BAC_{LSOCV} = 67.7% and AUC = 0.76 for the environmental plus clinical model; ROD: BAC_{LSOCV} = 70.5% and AUC = 0.81 for the clinical model, BAC_{LSOCV} = 62.5% and AUC = 0.72 for the environmental plus clinical model). Notably, the model integrating environmental, clinical and sMRI predictions, which achieved the best BAC_{LSOCV} in the CHR discovery sample, achieved much lower BAC (58.7%) when applied to the CHR replication sample, but the performance difference between CHR discovery and validation samples was not significant (P = 0.12; Supplementary Table 6). The transdiagnostic risk calculator built with the clinical data of the pooled CHR and ROD groups performed above chance in the pooled replication cohort (BAC_{LSOCV} = 69.7% and AUC = 0.74), in CHR alone (BAC_{LSOCV} = 61.1% and AUC = 0.75) and in ROD alone (BAC_{LSOCV} = 75% and AUC = 0.72) (Table 2 and Supplementary Table 4). Models combining clinical and environmental decision scores, and those combining clinical, environmental and sMRI decision scores, were those reaching the highest performance in all cohorts, and performed similarly (clinical + environmental model: BAC_{LSOCV} = 68.2% and AUC = 0.74 for CHR + ROD, BAC_{LSOCV} = 62.1% and AUC = 0.76 for CHR alone, and BAC_{LSOCV} = 71.5% and AUC = 0.73 for ROD alone; clinical + environmental + sMRI model: BAC_{LSOCV} = 69.6% and AUC = 0.75 for CHR + ROD, BAC_{LSOCV} = 67.4% and AUC = 0.77 for CHR alone, and BAC_{LSOCV} = 69.3% and AUC = 0.72 for ROD alone) (Table 2, Supplementary Tables 4 and 6).

Potential generalisation of risk calculators to relevant clinical trajectories

Linear mixed models results revealed that the prognostic role functioning assignments produced by the environmental plus clinical model stratified clinical trajectories of the ROD cohort, with respect to negative symptoms (P_{FDR} = 0.02) and environmental quality of life (P_{FDR} = 0.02; Fig. 3b). It did not stratify any clinical readout trajectory in the CHR group (all P_{FDR} > 0.4; Fig. 3a).

Environmental feature knock-out analysis

In the CHR group, the removal of one environmental variable at a time did not produce models superior to the original environmental classifier (Fig. 3, Supplementary Appendix 1, Section 8 and Supplementary Table 7), whereas in ROD, the model without the PAS childhood variable was superior to the original one (BAC_{LSOCV} = 60.1%). All other models performed similarly to the original one (Fig. 3d, Supplementary Appendix 1, Section 8 and Supplementary Table 7).

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

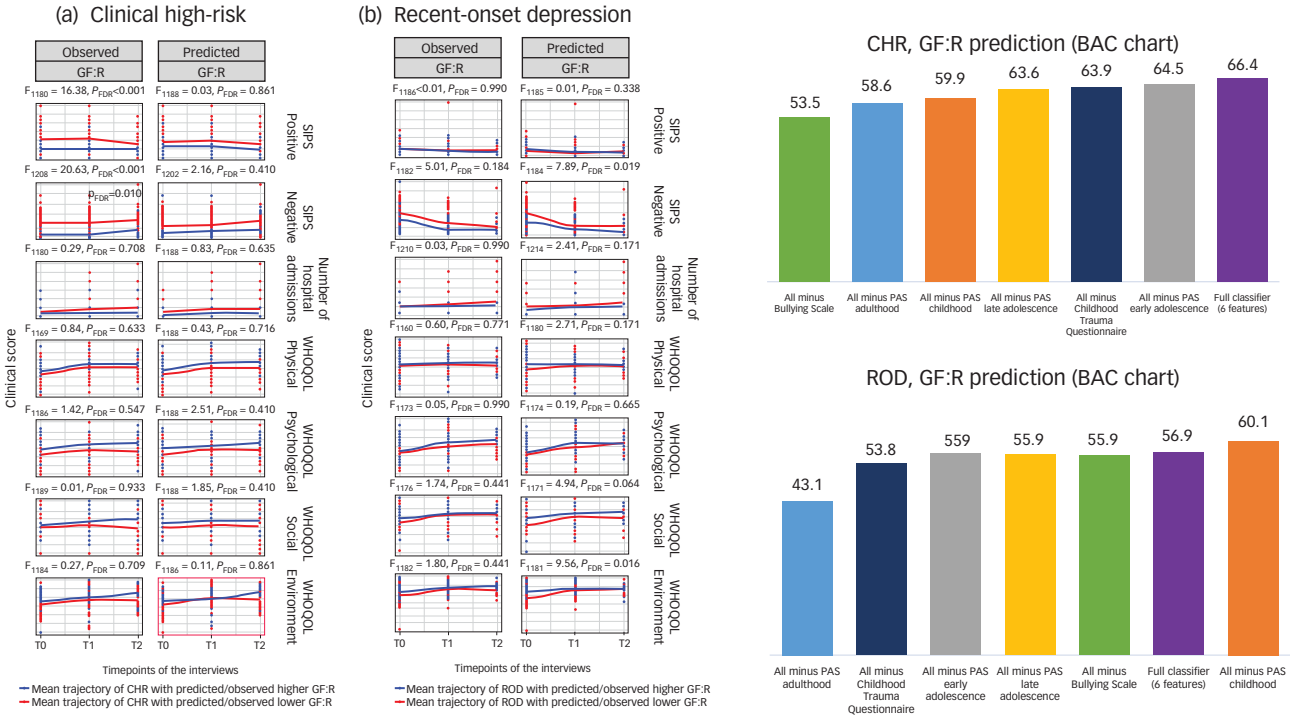

Fig. 3 Results of the linear mixed effects models used to compare main and slope differences between trajectories of prognostic or observed assignments (lower versus higher) for role functioning in CHR (a) and ROD (b). Linear fits are depicted in each plot. Significant main effects are highlighted in bold. Leave-site-out BAC percentage of all environmental models generated in CHR (c) and ROD (d) through recursive elimination of one environmental variable at a time. BAC, balanced accuracy; CHR, clinical high-risk; FDR, false discovery rate; GF:R, Global Functioning: Role; PAS, Premorbid Adjustment Scale; ROD, recent-onset depression; SIPS, Structured Interview for Psychosis-Risk Syndromes; WHOQOL, World Health Organization Quality of Life – Brief Questionnaire.

Investigation of between-classifiers relationships

In the CHR group, the sMRI-based model significantly predicted environmental decision scores, explaining 6.56% of the observed variance ($P_{FDR} = 0.02$; Supplementary Table 7). However, the clinical regression model could not predict environmental decision scores (explained variance 1.99%, $P_{FDR} = 0.24$). In the ROD group, the model predicting environmental-related decision scores for GF:R according to clinical data was significant ($P_{FDR} = 0.01$), and explained 9.18% of the observed variance (Supplementary Table 7). However, the sMRI regression model was non-significant (explained variance 0.03%, $P_{FDR} = 0.87$). Full performance metrics from each multivariate regression performed are reported in Supplementary Table 8.

Discussion

We demonstrated that, by combining environmental, clinical and sMRI baseline predictions, we could predict outcome in role functioning in CHR with 71.2% BAC_{LSOCV}, and significance, across seven geographically different European sites, but with much less accuracy and lower AUC than a CHR replication sample ($BAC_{LSOCV} = 58.7\%$). On the other hand, by combining clinical and environmental predictions without sMRI, we could predict outcome role functioning in CHR with 65.4% BAC_{LSOCV} in the discovery sample and 67.7% BAC_{LSOCV} in the replication sample, in both cases with an AUC of around 0.75. Therefore, our results support our hypothesis that environmental variables inform prediction of outcome in role functioning. Furthermore, they encourage future research to employ, set up or redefine machine learning algorithms for worse outcome prediction in a complex, superordinate and multimodal, rather than unimodal perspective.[7] Despite the good performance of our combined environmental and clinical risk calculator in CHR discovery and validation samples, linear mixed model analysis revealed that prognostic assignments (lower versus higher role outcome) were not associated with any other clinical trajectory. This finding suggests that the prognostic validity of this classifier seems limited to functional deficits and does not generalise to other clinical readouts. However, future studies involving the investigation of less state-affected clinical variables (i.e. academic functioning, work skills, resilience) should further clarify the extent of prognostic relevance of this model.

In ROD, only the combination of environmental and clinical variables predicted role outcome with significance and a modest BAC_{LSOCV} (58.9%), with an AUC of 0.60. It also pointed to a great extent of generalisability to the ROD replication sample ($BAC_{LSOCV} = 62.5\%$ and an even higher AUC of −0.72). Notably, despite lower BAC_{LSOCV} scores compared with CHR, linear mixed effects models used to calculate clinical trajectories in ROD revealed that role functioning outcome assignments based on environmental and clinical predictions stratified negative symptoms and environmental quality-of-life trajectories. This finding may indicate prognostic generalisation of this multimodal predictive model outside of its original role functioning domain.

Notably, accuracies in both discovery and replication samples were slightly lower in ROD than in CHR. A possible explanation emerges from the analysis of the prevalence of DSM-IV-TR diagnoses in both samples. A greater percentage of ROD individuals, relative to CHR, meet diagnostic criteria for at least one DSM-IV-TR psychiatric disorder. However, in our CHR samples, several and diverse DSM-IV-TR conditions are present, especially in mood and anxiety domains, consistent with previous literature.[56] Unlike

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

Table 3 Prevalence comparisons of the DSM-IV-TR diagnoses in the CHR and ROD discovery samples characterised by lower versus higher role functioning at baseline and follow-up examinations 9 months later

	T0									T1								
	CHR				ROD					CHR				ROD				
Characteristic	Lower GF:R	Higher GF:R	χ^2	*P*-value	Lower GF:R	Higher GF:R	χ^2	*P*-value	Characteristic	Lower GF:R	Higher GF:R	χ^2	*P*-value	Lower GF:R	Higher GF:R	χ^2	*P*-value
≥1 DSM-IV diagnosis: yes/no (% yes/% no)	45/7 (87/13)	29/11 (73/77)	2.83	0.09	48/0 (100/0)	45/2 (96/4)	2.09	0.15	≥1 DSM-IV diagnosis: yes/no (% yes/% no)	32/20 (62/38)	6/34 (15/85)	2.02	<0.01	32/16 (67/33)	19/28 (40/60)	6.58	0.01
Bipolar disorder type 1	1/51 (2/98)	0/40 (0/100)	0.78	0.38	0/48 (0/100)	0/47 (0/100)	–	–	Bipolar disorder type 1	1/51 (2/98)	0/40 (0/100)	0.78	0.38	0/48 (0/100)	0/47 (0/100)	–	–
Bipolar disorder type 2	0/52 (0/100)	2/38 (58/95)	2.66	0.10	0/48 (0/100)	0/47 (0/100)	–	–	Bipolar disorder type 2	0/52 (0/100)	0/40 (0/100)	–	–	0/48 (0/100)	0/47 (0/100)	–	–
Major depressive disorder	24/28 (46/53)	14/26 (35/65)	1.16	0.28	47/1 (98/2)	39/8 (83/17)	6.18	0.01	Major depressive disorder	14/38 (27/73)	0/40 (0/100)	1.27	<0.01	20/28 (42/58)	8/39 (17/83)	6.94	0.08
Dysthymic disorder	5/47 (10/90)	3/37 (7/93)	1.65	0.44	0/48 (0/100)	0/47 (0/100)	–	–	Dysthymic disorder	3/49 (6/94)	1/39 (2/98)	0.58	0.45	3/45 (6/94)	1/46 (2/98)	4.00	0.14
Panic disorder	8/44 (15/85)	1/39 (2/98)	4.25	0.04	4/44 (8/92)	1/46 (2/98)	1.83	0.18	Panic disorder	3/49 (6/94)	0/40 (0/100)	2.39	0.12	1/47 (2/98)	1/46 (2/98)	<0.01	0.99
Agoraphobia (AWOPD)	1/51 (2/98)	3/37 (7/93)	1.69	0.19	1/47 (2/98)	0/47 (0/100)	0.99	0.32	Agoraphobia (AWOPD)	2/50 (4/96)	1/39 (2/98)	0.13	0.72	0/48 (0/100)	2/45 (4/96)	2.09	0.15
Social phobia	12/40 (23/77)	2/38 (5/95)	5.73	0.02	4/44 (8/92)	3/44 (6/94)	0.13	0.72	Social phobia	2/50 (4/96)	1/39 (2/98)	0.13	0.72	2/46 (4/96)	3/44 (6/94)	0.23	0.63
Specific phobia	2/50 (4/96)	0/40 (0/100)	1.57	0.23	1/47 (2/98)	3/44 (6/94)	1.09	0.23	Specific phobia	0/52 (0/100)	0/40 (0/100)	1.31	0.25	0/48 (0/100)	4/43 (9/91)	4.26	0.04
Obsessive–compulsive disorder	6/46 (12/88)	3/37 (7/93)	0.42	0.52	1/47 (2/98)	1/46 (2/98)	<0.01	0.99	Obsessive–compulsive disorder	4/48 (8/92)	0/40 (0/100)	3.22	0.07	2/46 (4/96)	2/45 (4/96)	<0.01	0.98
Post-traumatic stress disorder	1/51 (2/98)	0/40 (0/100)	0.78	0.38	0/48 (0/100)	1/46 (2/98)	1.03	0.31	Post-traumatic stress disorder	0/52 (0/100)	0/40 (0/100)	–	–	0/48 (0/100)	0/47 (0/100)	–	–
Alcohol dependence	0/52 (0/100)	0/40 (0/100)	–	–	0/48 (0/100)	1/46 (2/98)	1.03	0.31	Alcohol dependence	3/49 (6/94)	0/40 (0/100)	2.39	0.12	0/48 (0/100)	0/47 (0/100)	–	–
Sedative-hypnotic-anxiolytic dependence	0/52 (0/100)	0/40 (0/100)	–	–	1/47 (2/98)	0/47 (0/100)	0.99	0.32	Sedative-hypnotic-anxiolytic dependence	0/52 (0/100)	0/40 (0/100)	–	–	0/48 (0/100)	0/47 (0/100)	–	–
Cannabis dependence	1/51 (2/98)	2/38 (4/96)	0.68	0.41	1/47 (2/98)	1/46 (2/98)	<0.01	0.99	Cannabis dependence	2/50 (4/96)	0/40 (0/100)	1.57	0.21	1/47 (2/98)	0/47 (0/100)	0.99	0.32

Analyses were performed for diagnoses in the domains of mood, anxiety and substance misuse. Presence of threshold diagnostic criteria in the past month before respective timepoint was examined, using χ^2-tests. For dysthymic disorder, lifetime presence of threshold and subthreshold criteria were combined and compared against absence of lifetime criteria. *P*-values were group- and timepoint-wise corrected for multiple comparisons, using the false discovery rate. Significance was defined at $\alpha = 0.05$. T0, baseline; T1, follow-up examinations 9 months after baseline; CHR, clinical high-risk; ROD, recent-onset depression; GF:R, Global Functioning Role; AWOPD, agoraphobia.

CHR, in ROD, the most represented DSM-IV-TR category across individuals is major depressive disorder. This pattern is observable in both discovery and replication samples, and across both time-points (Table 3 and Supplementary Table 2). However, this observed homogeneity of our ROD group is only partially consistent with previous literature, which indeed reported frequent comorbidities between depression and other psychiatric disorders.[57,58] We might speculate that the higher clinical variability in our CHR samples might have led to a more accurate and more representative multimodal predictive model for this population, and that the higher clinical homogeneity in our ROD samples may have driven the better generalisation of ROD models to other clinical trajectories. However, these hypotheses need to be tested by future studies; for example, studies using subtyping/clustering procedures.[59]

Taken together, our CHR and ROD findings suggest that individualised prediction of role functioning outcomes is possible, although with modest accuracy, in a replicable and geographically validated framework based on environmental and clinical information. sMRI, however, did not seem to play an important role in this prediction, as highlighted by the fact that the combined clinical-environmental-sMRI prediction model developed here could predict role outcome very well in the CHR discovery sample, but with much lower accuracy in the CHR validation sample, although the difference was marginally significant. This aligns with recent views cautioning researchers about neuroimaging-based machine learning findings, because of their excessively high dimensionality, especially in presence of small sample sizes and heterogeneous clinical phenomena.[60] However, a recently published study has highlighted that sMRI models show great predictive power when applied to transition-to-psychosis prediction.[61] Considering our findings and this recent evidence, we may speculate that structural neuroimaging-based information could be more informative for diagnostic prediction, rather than for transdiagnostic outcomes. Future studies are strongly warranted to validate this hypothesis.

Consistently, our assessment of transdiagnostic generalisability supported the prognostic relevance of the combination of environmental and clinical risk calculators for role prediction, showing accuracy (62.4% BAC_{LSOCV} and 0.71 AUC), significance and generalisability to any replication sample combination. Notably, also the combination of environmental, clinical and sMRI decision scores led to accurate (61.1% BAC_{LSOCV} and 0.67 AUC), significant and generalisable findings, and its performance metrics were very similar to those of the combined clinical and environmental risk calculator (Table 2 and Supplementary Table 6). These findings further support the view that MRI data carry a negligible amount of predictive information when applied to the longitudinal investigation of transdiagnostic outcomes.

Interestingly, investigation of the reliability of features within predictive models built on both separate and pooled on CHR and ROD samples revealed that premorbid adjustment in adulthood was transdiagnostically important for outcome prediction. This suggests that this environmental feature might be associated with role functional outcome regardless of the clinical population tested, and that the transdiagnostic potential of our combined environmental and clinical risk calculator is mainly driven by this environmental feature. Notably, the PAS scale measures the degree of achievement of developmental goals over time, according to gender, socioeconomic status and age.[62] Therefore, adult maladjustments to developmental goals may be more predictive compared with other features, because they represent the most proximal-to-baseline environmental feature in the algorithm, and may indeed be a result of earlier environmental adverse events.[14,46] Consistently, besides premorbid adjustment in adulthood, the combinations of environmental adverse events with the highest predictive value were different between CHR- and ROD-based risk calculators, but all included the presence of earlier developmental maladjustment. Indeed, after PAS adulthood, the most predictive feature for role functioning prediction in the CHR sample was the occurrence of childhood trauma,[39] and, in ROD, early adolescence maladjustment to environment.[63] Findings therefore highlight the existence of both transdiagnostic and syndrome-specific environmental adverse events able to predict role outcome in different clinical populations.[64] Notably, environmental adverse events occurred in different time periods may increase the risk for disease in a composite/compounding way.[65] For example, it may be hypothesised that childhood adversities (e.g. childhood trauma) may increase the risk for subsequent maladjustment (e.g. lower levels of environmental adjustment in adulthood) by increasing the risk of exposure to further environmental stressors (e.g. bullying victimisation), thus acting as triggers of a causal environmental path.[66] However, childhood adversities may either predict subsequent adversities, or interact with other adult adversities,[67] thus making the picture even more complex. This view is consistent with findings from our recursive feature elimination procedure, which showed that the most accurate environmental model in CHR was the one constructed on all environmental adverse events, hence reiterating the importance of taking into account the complex *gestalt* of environmental maladjustments and adverse events.

Finally, we observed that environmental decision scores predicting follow-up role outcome were significantly associated in CHR, with sMRI baseline data, and in ROD, with clinical data. These preliminary findings partially support the hypothesis that the history of environmental adverse events may mediate differential associations of baseline clinical and sMRI predictors with follow-up role outcome. However, future path analysis studies investigating the specific mediating or moderating role of environmental events in the relationship between clinical data, GMVs and follow-up outcome are warranted.

Limitations

This study has some limitations. Although our study was based on an extension of previously published prediction models,[7] we could not directly compare the two sets of results, as the samples did not exactly overlap. This was because complete lack of environmental assessments for some of the patients in the CHR and ROD samples. Furthermore, it should be noted that the small sample size did not allow us to investigate potential gender effects on role functioning predictions. As gender is differentially linked to environmental adversity effects,[68] future studies should further investigate this relationship. Also, calibration results (Fig. 1 and Supplementary Appendix 1, Section 10) revealed relatively low, although not perfect, ECEs. This seems to be frequent with Support Vector Machine algorithms.[69] Future studies might take into account calibration already the model-building phase, e.g. *via* Bayesian Binning into Quantiles, to try to achieve better calibration performance.[70] Moreover, the type of environmental information collected might represent a limit. As well as developmental maladjustments, childhood traumatic experiences and bullying victimisation have been all previously associated with psychosis[39,41] and depression;[43,44,46] information regarding other adverse life events that are known risk factors for psychosis and other mental disorders were not collected within this study.[71] Furthermore, our environmental predictive model was based on variables reflecting the occurrence of environmental adverse events, as well as variables reflecting the level of environmental adjustment to such adverse events (and many others), as measured by PAS. It cannot be excluded, therefore, that PAS score variations may be the result of not only the occurrence of environmental adversities across the lifespan, but also individual differences in the ability of adapting to adverse environmental exposures, thus having a close relationship with core adjustment-related psychological attributes, such as

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

coping strategies and resilience.[72] This relationship needs to be thoroughly and experimentally investigated by future studies.

Importantly, it should be noted that our validation sample was recruited within the same study of the discovery sample, although a part of the validation sample was recruited at different sites to the discovery ones. Furthermore, we employed the NeuroMiner software to carry out our machine learning pipeline, as one of the main aims of the PRONIA consortium was to facilitate open science and validation of findings *via* the NeuroMiner Model Library (see Data availability). However, the current NeuroMiner version allowed us to perform permutation testing only on the discovery cohorts, without providing significance estimates for replication performance. Future validation of the model in completely independent populations from other consortia and countries is warranted, to better account for optimism in the performance estimate and further characterise the performance of our models. Another important limitation of our study is the small sample size. CHR is a difficult population to recruit and keep in a longitudinal study, because of the risk-related aspect of their condition. This issue is quite common; indeed, our sample size is in line with other samples employed in recent machine learning studies conducted on CHR populations.[73,74] Moreover, in our case, it should be noted that findings are further limited by the number of features employed in the models. Machine learning-based predictions require a large amount of data,[75] but the number of features we used was limited. Consistently, within our machine learning pipeline, the number of events per predictor variable is lower than recently recommended.[76] However, this is not uncommon in the CHR field. Indeed, although a large number of input features reduces the risk of overoptimistic results, it could be difficult to translate models based on a large number of features into clinical, real-world settings, where data obtained are usually limited because of patient adherence, drop out or time constraints. With this regard, it has been previously suggested[77] to employ double-cycle, nested, leave-site-out cross-validation techniques as a gold-standard strategy to mitigate overfitting and optimism of models' performance, especially in cases of limited numbers of features and/or of individuals, as in this study. Nevertheless, taking all of these limitations into account, it should be noted that the understanding of the clinical usefulness of the presented models is strictly dependent on the results of further validation on larger and geographically diverse cohorts. Future studies are warranted to thoroughly investigate the stability and generalisability of our models, and further test the potential of translation into clinical practice of our models. With respect to the clinical implementation of our models, it should be also noted that the threshold we have chosen for the classification metrics was not determined based on clinical grounds. Indeed, CHR and ROD are very heterogeneous clinical states, and obtaining consensus-based risk estimates for these help-seeking populations, especially for transdiagnostic outcomes like occupational functioning, is a challenge. An online machine learning-based strategy has been recently proposed in a publication aiming at developing psychosis predictive models in diverse at-risk populations across different consortia.[25] Future studies are warranted to test the feasibility of such solution, and to direct efforts toward the generation of a public library of machine learning-based estimated risk distributions reflecting diverse help-seeking populations.

In conclusion, we explored syndrome-specific and transdiagnostic predictive models combining clinical and environmental information, which could predict role outcome with moderate accuracy in several independent samples. For ROD, these predictions seem to be prognostically relevant to non-functioning clinical trajectories, like negative symptoms and quality of life. The modest, although stable across samples, performance of our combined clinical and environmental predictive models (both the syndrome-specific and transdiagnostic) encourage future research to spend significant efforts in further validating existing multimodal risk calculators to fully assess their degree of applicability in healthcare settings, as well as in defining guidelines for models' comparability and replicability.[26] If geographically and extensively validated, risk calculators built on both CHR and affective populations could benefit patients with a realistic, personalised prognostic estimation of their functioning level irrespective of diagnostic boundaries, thus supporting early rehabilitation and a better integration of patients into their societal environment. However, future studies on environmental adverse events are warranted to define to what extent risk factors interact with each other, with symptoms profiles and with neurobiological alterations, to predispose young individuals for worse role outcomes. Such multimodal frameworks will more likely mirror the complex and heterogeneous architecture of psychosis and depression risk, and would hopefully contribute to provide models with even higher accuracies, closer to real-world scenarios, and with more potential for translation into clinical practice.

Linda A. Antonucci, Department of Education Science, Psychology and Communication Science, University of Bari Aldo Moro, Italy; and Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; **Nora Penzel**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; and Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; **Rachele Sanfelici**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; and Institute for Psychiatry, Max Planck School of Cognition, Germany; **Alessandro Pigoni**, Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Italy; and Social and Affective Neuroscience Group, MoMiLab, IMT School for Advanced Studies Lucca, Italy; **Lana Kambeitz-Ilankovic**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; and Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; **Dominic Dwyer**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; **Anne Ruef**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; **Mark Sen Dong**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; **Ömer Faruk Öztürk**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; and Institute for Psychiatry, International Max Planck Research School for Translational Psychiatry, Germany; **Katharine Chisholm**, Institute for Mental Health, University of Birmingham, UK; and Department of Psychology, Aston University, UK; **Theresa Haidl**, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; **Marlene Rosen**, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; **Adele Ferro**, Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Italy; **Giulio Pergola**, Department of Basic Medical Science, Neuroscience, and Sense Organs, University of Bari Aldo Moro, Italy; **Ileana Andriola**, Department of Basic Medical Science, Neuroscience, and Sense Organs, University of Bari Aldo Moro, Italy; **Giuseppe Blasi**, Department of Basic Medical Science, Neuroscience, and Sense Organs, University of Bari Aldo Moro, Italy; **Stephan Ruhrmann**, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; **Frauke Schultze-Lutter**, Department of Psychiatry and Psychotherapy, Heinrich-Heine University Düsseldorf, Germany; Department of Psychology and Mental Health, Faculty of Psychology, Airlangga University, Indonesia; and University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Switzerland; **Peter Falkai**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; **Joseph Kambeitz**, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; **Rebekka Lencer**, Institute for Translational Psychiatry, University of Münster, UK; and Department of Psychiatry and Psychotherapy, University of Lübeck, Germany; **Udo Dannlowski**, Institute for Translational Psychiatry, University of Münster, UK; **Rachel Upthegrove**, Institute for Mental Health, University of Birmingham, UK; and Early Intervention Service, Birmingham Women's and Children's NHS Foundation Trust, UK; **Raimo K. R. Salokangas**, Department of Psychiatry, University of Turku, UK; **Christos Pantelis**, Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne, Australia; **Eva Meisenzahl**, Department of Psychiatry and Psychotherapy, Heinrich-Heine University Düsseldorf, Germany; **Stephen J. Wood**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; Orygen, Australia; Centre for Youth Mental Health, University of Melbourne, Australia; and School of Psychology, University of Birmingham, UK; **Paolo Brambilla**, Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Italy; and Department of Pathophysiology and Transplantation, University of Milan, Italy; **Stefan Borgwardt**, Institute for Translational Psychiatry, University of Münster, UK; and Department of Psychiatry (Psychiatric University Hospital, University Psychiatric Clinics Basel), University of Basel, Switzerland; **Alessandro Bertolino**, Department of Basic Medical Science, Neuroscience, and Sense Organs, University of Bari Aldo Moro, Italy; **Nikolaos Koutsouleris**, Department of Psychiatry and Psychotherapy, Ludwig Maximilians University Munich, Germany; **the PRONIA Consortium**

Correspondence: Linda A. Antonucci. Email: linda.antonucci@uniba.it

First received 20 May 2021, final revision 8 Nov 2021, accepted 17 Jan 2022

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1192/bjp.2022.16

Data availability

The combined clinical and environmental prediction models can be found in the NeuroMiner Model Library (http://www.proniapredictors.eu/#). All analysis pipelines from NeuroMiner are publicly available via GitHub (https://github.com/neurominer-git/).

Acknowledgements

The PRONIA consortium:

The authors listed here performed the screening, recruitment, rating, examination and follow-up of the study participants. They were involved in implementing the examination protocols of the study, setting up its IT infrastructure, and organising the flow and quality control of the data analysed in this manuscript between the local study sites and the central study database.

Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Munich, Germany: Shalaila Haas, Alkomiet Hasan, Claudius Hoff, Ifrah Khanyaree, Aylin Melo, Susanna Muckenhuber-Sternbauer, Yanis Köhler, Ömer Öztürk, Nora Penzel, David Popovic, Adrian Rangnick, Sebastian von Saldern, Rachele Sanfelici, Moritz Spangemacher, Ana Tupac, Maria Fernanda Urquijo, Johanna Weiske, Antonia Wosgien, Camilla Krämer.

Department of Psychiatry and Psychotherapy, University Hospital and Faculty of Medicine, University of Cologne, Cologne, Germany: Karsten Blume, Dominika Julkowski, Nathalie Kaden, Ruth Milz, Alexandra Nikolaides, Mauro Seves, Silke Vent, Martina Wassen.

Department of Psychiatry (Psychiatric University Hospital, University Psychiatric Clinics Basel), University of Basel, Switzerland: Christina Andreou, Laura Egloff, Fabienne Harrisberger, Ulrike Heitz, Claudia Lenz, Letizia Leanza, Amatya Mackintosh, Renata Smieskova, Erich Studerus, Anna Walter, Sonja Widmayer.

Institute of Mental Health & School of Psychology, University of Birmingham, UK: Chris Day, Sian Lowri Griffiths, Mariam Iqbal, Paris Lalousis, Mirabel Pelton, Pavan Mallikarjun, Alexandra Stainton, Ashleigh Lin.

Department of Psychiatry, University of Turku, Finland: Alexander Denissoff, Anu Ellilä, Tiina From, Markus Heinimaa, Tuula Ilonen, Päivi Jalo, Heikki Laurikainen, Antti Luutonen, Akseli Mäkela, Janina Paju, Henri Pesonen, Reetta-Liina Säilä, Anna Toivonen, Otto Turtonen.

Department of Psychiatry (Psychiatric University Hospital LVR/HHU Düsseldorf), University of Düsseldorf, Germany: Sonja Botterweck, Norman Kluthausen, Gerald Antoch, Julian Caspers, Hans-Jörg Wittsack.

General Electric Global Research Inc., USA: Ana Beatriz Solana, Manuela Abraham, Timo Schirmer.

Workgroup of Paolo Brambilla, University of Milan, Italy:

Department of Neuroscience and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Italy: Marika Belleri, Francesca Bottinelli, Giuseppe Delvecchio, Adele Ferro, Eleonora Maggioni, Marta Re, Letizia Squarcina; Programma2000, Niguarda Hospital, Italy: Emiliano Monzani, Maurizio Sberna; San Paolo Hospital, Italy: Armando D'Agostino, Lorenzo Del Fabro; Villa San Benedetto Menni, Albese con Cassano: Giampaolo Perna, Maria Nobile, Alessandra Alciati.

Workgroup of Paolo Brambilla at the University of Udine, Italy:

Department of Medical Area, University of Udine, Udine, Italy: Matteo Balestrieri, Carolina Bonivento, Giuseppe Cabras, Franco Fabbro.

Author contributions

N.K. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. L.A.A. and N.K. conceived and designed the study. N.K., R.S. A.P., N.P., L.K.-I., S.R., A.R., D.D., K.C., J.K., T.H., F.S.-L., C.P., S.J.W., P.B., S.B., A.B., R.K.R.S., E.M., R.U., A.F., M.R., I.A., O.F.O. and M.S.D. were responsible for data acquisition, analysis or interpretation. L.A.A., R.S. and N.K. drafted the manuscript. N.P., R.S., G.P., G.B, F.S.-L., T.H., A.P., R.L., A.F., K.C., U.D. and R.U. critically revised the manuscript for important intellectual content. N.K., L.K.-I., E.M., S.R., R.K.R.S., C.P., P.B., S.B., S.J.W., A.B. and R.U. obtained funding. N.K., P.F. and A.B. supervised the study.

Funding

This work was supported by the EU-FP7-HEALTH grant for the project 'PRONIA' (Personalized Prognostic Tools for Early Psychosis Management; PI: N.K., agreement number: 602152) and by the Structural European Funding of the Italian Minister of Education and Research (Attraction and International Mobility – AIM - action, grant agreement No 1859959). The AIM action also funds L.A.A.'s salary.

Declaration of interest

A.B. has received lecture fees from Otsuka, Janssen and Lundbeck; and consultant fees from Biogen. N.K. has received honoraria for talks presented at education meetings organised by Otsuka/Lundbeck. N.K. and E.M. hold commercial patents that are related to the present work (https://patents.google.com/patent/US20160192889/). C.P. participated in advisory boards for Janssen-Cilag, AstraZeneca, Lundbeck and Servier; received honoraria for talks presented at educational meetings organized by AstraZeneca, Janssen-Cilag, Eli Lilly, Pfizer, Lundbeck and Shire; and was supported by National Health and Medical Research Council Senior Principal Research Fellowship (grants 628386 and 1105825) and European Union–National Health and Medical Research Council (grant 1075379). G.P.'s position is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 798181. L.A.A.'s salary is funded from the Structural European Funding of the Italian Minister of Education (Attraction and International Mobility – AIM - action, grant agreement number 1859959).

L.K.-I., D.D., F.S.-L. and R.U. are members of the *BJPsych* editorial board and did not take part in the review or decision-making process of this paper. No other disclosures were reported.

References

1 Bora E, Harrison BJ, Yucel M, Pantelis C. Cognitive impairment in euthymic major depressive disorder: a meta-analysis. *Psychol Med* 2013; **43**(10): 2017–26.

2 Ruhrmann S, Paruch J, Bechdolf A, Pukrop R, Wagner M, Berning J, et al. Reduced subjective quality of life in persons at risk for psychosis. *Acta Psychiatr Scand* 2008; **117**(5): 357–68.

3 Schultze-Lutter F, Michel C, Ruhrmann S, Schimmelmann BG. Prevalence and clinical relevance of interview-assessed psychosis-risk symptoms in the young adult community. *Psychol Med* 2018; **48**(7): 1167–78.

4 Harvey PD, Strassnig M. Predicting the severity of everyday functional disability in people with schizophrenia: cognitive deficits, functional capacity, symptoms, and health status. *World Psychiatry* 2012; **11**(2): 73–9.

5 Yung AR, Yuen HP, McGorry PD, Phillips LJ, Kelly D, Dell'Olio M, et al. Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. *Aust N Z J Psychiatry* 2005; **39**(11–12): 964–71.

6 Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol* 2018; **14**: 91–118.

7 Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, et al. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry* 2018; **75**(11): 1156–72.

8 Burton CZ, Tso IF, Carrion RE, Niendam T, Adelsheim S, Auther AM, et al. Baseline psychopathology and relationship to longitudinal functional outcome in attenuated and early first episode psychosis. *Schizophr Res* 2019; **212**: 157–62.

9 Carrion RE, Goldberg TE, McLaughlin D, Auther AM, Correll CU, Cornblatt BA. Impact of neurocognition on social and role functioning in individuals at clinical high risk for psychosis. *Am J Psychiatry* 2011; **168**(8): 806–13.

10 Carrion RE, Auther AM, McLaughlin D, Addington J, Bearden CE, Cadenhead KS, et al. Social decline in the psychosis prodrome: predictor potential and heterogeneity of outcome. *Schizophr Res* 2021; **227**: 44–51.

11 Velthorst E, Zinberg J, Addington J, Cadenhead KS, Cannon TD, Carrion RE, et al. Potentially important periods of change in the development of social and role functioning in youth at clinical high risk for psychosis. *Dev Psychopathol* 2018; **30**(1): 39–47.

12 Cornblatt BA, Carrion RE, Addington J, Seidman L, Walker EF, Cannon TD, et al. Risk factors for psychosis: impaired social and role functioning. *Schizophr Bull* 2012; **38**(6): 1247–57.

13 Evert H, Harvey C, Trauer T, Herrman H. The relationship between social networks and occupational and self-care functioning in people with psychosis. *Soc Psychiatry Psychiatr Epidemiol* 2003; **38**(4): 180–8.

14 Howes OD, Murray RM. Schizophrenia: an integrated sociodevelopmental-cognitive model. *Lancet* 2014; **383**(9929): 1677–87.

15 Kwong ASF, Lopez-Lopez JA, Hammerton G, Manley D, Timpson NJ, Leckie G, et al. Genetic and environmental risk factors associated with trajectories of depression symptoms from adolescence to young adulthood. *JAMA Netw Open* 2019; **2**(6): e196587.

16 Schmitt A, Malchow B, Hasan A, Falkai P. The impact of environmental factors in severe psychiatric disorders. *Front Neurosci* 2014; **8**: 19.

17 Baker LM, Williams LM, Korgaonkar MS, Cohen RA, Heaps JM, Paul RH. Impact of early vs. late childhood early life stress on brain morphometrics. *Brain Imaging Behav* 2013; **7**(2): 196–203.

18 Carballedo A, Lisiecka D, Fagan A, Saleh K, Ferguson Y, Connolly G, et al. Early life adversity is associated with brain changes in subjects at family risk for depression. *World J Biol Psychiatry* 2012; **13**(8): 569–78.

19 Popovic D, Ruef A, Dwyer DB, Antonucci LA, Eder J, Sanfelici R, et al. Traces of trauma: a multivariate pattern analysis of childhood trauma, brain structure, and clinical phenotypes. *Biol Psychiatry* 2020; **88**(11): 829–42.

20 Schultze-Lutter F, Schimmelmann BG, Michel C. Clinical high-risk of and conversion to psychosis in the community: a 3-year follow-up of a cohort study. *Schizophr Res* 2021; **228**: 616–8.

21 Fusar-Poli P, Nelson B, Valmaggia L, Yung AR, McGuire PK. Comorbid depressive and anxiety disorders in 509 individuals with an at-risk mental state: impact on psychopathology and transition to psychosis. *Schizophr Bull* 2014; **40**(1): 120–31.

22 Fusar-Poli P, Salazar de Pablo G, Correll CU, Meyer-Lindenberg A, Millan MJ, Borgwardt S, et al. Prevention of psychosis: advances in detection, prognosis, and intervention. *JAMA Psychiatry* 2020; **77**(7): 755–65.

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

23 Fusar-Poli P, Tantardini M, De Simone S, Ramella-Cravaro V, Oliver D, Kingdon J, et al. Deconstructing vulnerability for psychosis: meta-analysis of environmental risk factors for psychosis in subjects at ultra high-risk. *Eur Psychiatry* 2017; **40**: 65–75.

24 Lee TY, Lee J, Kim M, Choe E, Kwon JS. Can we predict psychosis outside the clinical high-risk state? A systematic review of non-psychotic risk syndromes for mental disorders. *Schizophr Bull* 2018; **44**(2): 276–85.

25 Koutsouleris N, Worthington M, Dwyer DB, Kambeitz-Ilankovic L, Sanfelici R, Fusar-Poli P, et al. Toward generalizable and transdiagnostic tools for psychosis prediction: an independent validation and improvement of the NAPLS-2 risk calculator in the multisite PRONIA cohort. *Biol Psychiatry* 2021; **90** (9): 632–42.

26 Rosen M, Betz LT, Schultze-Lutter F, Chisholm K, Haidl TK, Kambeitz-Ilankovic L, et al. Towards clinical application of prediction models for transition to psychosis: a systematic review and external validation study in the PRONIA sample. *Neurosci Biobehav Rev* 2021; **125**: 478–92.

27 Antonucci LA, Pergola G, Pigoni A, Dwyer D, Kambeitz-Ilankovic L, Penzel N, et al. A pattern of cognitive deficits stratified for genetic and environmental risk reliably classifies patients with schizophrenia from healthy control subjects. *Biological Psychiatry* 2019; **87**: 697–707.

28 Cannon TD, Yu C, Addington J, Bearden CE, Cadenhead KS, Cornblatt BA, et al. An individualized risk calculator for research in prodromal psychosis. *Am J Psychiatry* 2016; **173**(10): 980–8.

29 Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 2016; **3**(3): 243–50.

30 Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Cai T, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol Psychiatry* 2016; **21**(10): 1366–71.

31 Grassi M, Perna G, Caldirola D, Schruers K, Duara R, Loewenstein DA. A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion in individuals with mild and premild cognitive impairment. *J Alzheimers Dis* 2018; **61**(4): 1555–73.

32 Feng R, Badgeley M, Mocco J, Oermann EK. Deep learning guided stroke management: a review of clinical applications. *J Neurointerv Surg* 2018; **10**(4): 358–62.

33 Bodatsch M, Ruhrmann S, Wagner M, Muller R, Schultze-Lutter F, Frommann I, et al. Prediction of psychosis by mismatch negativity. *Biol Psychiatry* 2011; **69**(10): 959–66.

34 Koutsouleris N, Wobrock T, Guse B, Langguth B, Landgrebe M, Eichhammer P, et al. Predicting response to repetitive transcranial magnetic stimulation in patients with schizophrenia using structural magnetic resonance imaging: a multisite machine learning analysis. *Schizophr Bull* 2018; **44**(5): 1021–34.

35 Koutsouleris N, Upthegrove R, Wood SJ. Importance of variable selection in multimodal prediction models in patients at clinical high risk for psychosis and recent onset depression-reply. *JAMA Psychiatry* 2019; **76**(3): 339–40.

36 Koutsouleris N, Kahn RS, Chekroud AM, Leucht S, Falkai P, Wobrock T, et al. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *Lancet Psychiatry* 2016; **3**(10): 935–46.

37 Kambeitz-Ilankovic L, Meisenzahl EM, Cabral C, von Saldern S, Kambeitz J, Falkai P, et al. Prediction of outcome in the psychosis prodrome using neuroanatomical pattern classification. *Schizophr Res* 2016; **173**(3): 159–65.

38 Lin A, Wood SJ, Nelson B, Beavan A, McGorry P, Yung AR. Outcomes of nontransitioned cases in a sample at ultra-high risk for psychosis. *Am J Psychiatry* 2015; **172**(3): 249–58.

39 Loewy RL, Corey S, Amirfathi F, Dabit S, Fulford D, Pearson R, et al. Childhood trauma and clinical high risk for psychosis. *Schizophr Res* 2019; **205**: 10–4.

40 Pergola G, Papalino M, Gelao B, Sportelli L, Vollerbergh W, Grattagliano I, et al. Evocative gene-environment correlation between genetic risk for schizophrenia and bullying victimization. *World Psychiatry* 2019; **18**(3): 366–7.

41 Tarbox SI, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA, Perkins DO, et al. Premorbid functional development and conversion to psychosis in clinical high-risk youths. *Dev Psychopathol* 2013; **25**(4 Pt 1): 1171–86.

42 Upthegrove R. Bullying, victimisation, and psychosis. *Lancet Psychiatry* 2015; **2**(7): 574–6.

43 Hill RM, Mellick W, Temple JR, Sharp C. The role of bullying in depressive symptoms from adolescence to emerging adulthood: a growth mixture model. *J Affect Disord* 2017; **207**: 1–8.

44 Negele A, Kaufhold J, Kallenbach L, Leuzinger-Bohleber M. Childhood trauma and its relation to chronic depression in adulthood. *Depress Res Treat* 2015; **2015**: 650804.

45 Scher CD, Stein MB, Asmundson GJ, McCreary DR, Forde DR. The childhood trauma questionnaire in a community sample: psychometric properties and normative data. *J Trauma Stress* 2001; **14**(4): 843–57.

46 Vocisano C, Klein DN, Keefe RS, Dienst ER, Kincaid MM. Demographics, family history, premorbid functioning, developmental characteristics, and course of patients with deteriorated affective disorder. *Am J Psychiatry* 1996; **153**(2): 248–55.

47 Cornblatt BA, Auther AM, Niendam T, Smith CW, Zinberg J, Bearden CE, et al. Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophr Bull* 2007; **33** (3): 688–702.

48 Haidl T, Schneider N, Dickmann K, Ruhrmann S, Kaiser N, Rosen M, et al. Validation of the Bullying Scale for Adults - results of the PRONIA-study. *J Psychiatr Res* 2020; **129**: 88–97.

49 Shapiro DI, Marenco S, Spoor EH, Egan MF, Weinberger DR, Gold JM. The Premorbid Adjustment Scale as a measure of developmental compromise in patients with schizophrenia and their healthy siblings. *Schizophr Res* 2009; **112**(1–3): 136–42.

50 Garcia D, Al Nima A, Kjell ON. The affective profiles, psychological well-being, and harmony: environmental mastery and self-acceptance predict the sense of a harmonious life. *PeerJ* 2014; **2**: e259.

51 Rhemtulla M, Brosseau-Liard PE, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods* 2012; **17**(3): 354–73.

52 Wolpert DH. Stacked generalization. *Neural Networks* 1992; **5**: 241–59.

53 Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning (Sydney, Australia, 2017)*. PMLR 70, 2017.

54 Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Cannon T, Ventura J, et al. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophr Bull* 2003; **29**(4): 703–15.

55 Skevington SM, Lotfy M, O'Connell KA, Group W. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual Life Res* 2004; **13**(2): 299–310.

56 Woods SW, Powers AR 3rd, Taylor JH, Davidson CA, Johannesen JK, Addington J, et al. Lack of diagnostic pluripotentiality in patients at clinical high risk for psychosis: specificity of comorbidity persistence and search for pluripotential subgroups. *Schizophr Bull* 2018; **44**(2): 254–63.

57 Thaipisuttikul P, Ittasakul P, Waleeprakhon P, Wisajun P, Jullagate S. Psychiatric comorbidities in patients with major depressive disorder. *Neuropsychiatr Dis Treat* 2014; **10**: 2097–103.

58 Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry* 2018; **75**(4): 336–46.

59 Dwyer DB, Cabral C, Kambeitz-Ilankovic L, Sanfelici R, Kambeitz J, Calhoun V, et al. Brain subtyping enhances the neuroanatomical discrimination of schizophrenia. *Schizophr Bull* 2018; **44**(5): 1060–9.

60 Vieira S, Gong Q, Scarpazza C, Lui S, Huang X, Crespo-Facorro B, et al. Neuroanatomical abnormalities in first-episode psychosis across independent samples: a multi-centre mega-analysis. *Psychol Med* 2021; **51**(2): 340–50.

61 Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R, et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry* 2021; **78**(2): 195–209.

62 Cannon-Spoor HE, Potkin SG, Wyatt RJ. Measurement of premorbid adjustment in chronic schizophrenia. *Schizophr Bull* 1982; **8**(3): 470–84.

63 Tyborowska A, Volman I, Niermann HCM, Pouwels JL, Smeekens S, Cillessen AHN, et al. Early-life and pubertal stress differentially modulate grey matter development in human adolescents. *Sci Rep* 2018; **8**(1): 9201.

64 Oliver D, Radua J, Reichenberg A, Uher R, Fusar-Poli P. Psychosis Polyrisk Score (PPS) for the detection of individuals at-risk and the prediction of their outcomes. *Front Psychiatry* 2019; **10**: 174.

65 Bhavsar V, Boydell J, McGuire P, Harris V, Hotopf M, Hatch SL, et al. Childhood abuse and psychotic experiences - evidence for mediation by adulthood adverse life events. *Epidemiol Psychiatr Sci* 2019; **28**(3): 300–9.

66 Morgan C, Reininghaus U, Fearon P, Hutchinson G, Morgan K, Dazzan P, et al. Modelling the interplay between childhood and adult adversity in pathways to psychosis: initial evidence from the AESOP study. *Psychol Med* 2014; **44**(2): 407–19.

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press

67 Hafeman DM, Schwartz S. Opening the black box: a motivation for the assessment of mediation. *Int J Epidemiol* 2009; **38**(3): 838–45.

68 Evans EA, Grella CE, Upchurch DM. Gender differences in the effects of childhood adversity on alcohol, drug, and polysubstance-related disorders. *Soc Psychiatry Psychiatr Epidemiol* 2017; **52**(7): 901–12.

69 Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc* 2020; **27**(4): 621–33.

70 Naeini MP, Cooper GF, Hauskrecht M. Obtaining well calibrated probabilities using Bayesian binning. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (Austin, Texas, 25–30 Jan 2015).* Association for the Advancement of Artificial Intelligence, 2015.

71 Stilo SA, Murray RM. Non-genetic factors in schizophrenia. *Curr Psychiatry Rep* 2019; **21**(10): 100.

72 Holz NE, Tost H, Meyer-Lindenberg A. Resilience and the brain: a key role for regulatory circuits linked to social stress and support. *Mol Psychiatry* 2020; **25**(2): 379–96.

73 Haining K, Brunner G, Gajwani R, Gross J, Gumley AI, Lawrie SM, et al. The relationship between cognitive deficits and impaired short-term functional outcome in clinical high-risk for psychosis participants: a machine learning and modelling approach. *Schizophr Res* 2021; **231**: 24–31.

74 Mongan D, Focking M, Healy C, Susai SR, Heurich M, Wynne K, et al. Development of proteomic prediction models for transition to psychotic disorder in the clinical high-risk state and psychotic experiences in adolescence. *JAMA Psychiatry* 2021; **78**(1): 77–90.

75 van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014; **14**: 137.

76 Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE Jr, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: part I - continuous outcomes. *Stat Med* 2019; **38**(7): 1262–75.

77 Sanfelici R, Dwyer DB, Antonucci LA, Koutsouleris N. Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: a meta-analytic view on the state of the art. *Biol Psychiatry* 2020; **88**: 349–60.

EXTRA CONTENT ONLINE

Extra

Auschwitz: 2. Children

Greg Wilkinson

Nazi Germany and its collaborators killed about 1.5 million Jewish children and tens of thousands of 'Gypsy' children, 5000–7000 German children with physical and mental disabilities living in institutions, as well as many Polish children and children residing in the German-occupied Soviet Union: adolescents had a greater chance of survival – in forced labour.[a]

Upon arrival at Auschwitz-Birkenau and other camps, the vast majority of young Jewish children were sent directly to the gas chambers. According to Dr Janina Kościuszkowa / 36319 (1897–1974), the children of Auschwitz concentration camp were divided into four groups: burned to death immediately on arrival; killed in their mothers' wombs or as soon as they were born; born in the camp and allowed to live; deported to the camp as prisoners.

There appear to be few medical accounts of children's mental states in Auschwitz. Robert Waitz / 157261 (1900–1978), professor in the faculty of medicine at Strasbourg, contributed two contrasting portraits[b] in *Témoignages Strasbourgeois* (1947).

> As can be seen in the adults, some children collapse morally and present themselves dirty, very pale, with a vague and anxious look, hardly responding to questions asked of them. Others, on the contrary, keep their equilibrium, stay clean, polite, affectionate.
>
> May I be allowed to give examples of two children's behaviour.
>
> One of these children is a young Luxemburger, arrived aged 13½ years, with his father and his brothers from Treblinka, one of the extermination camps in the Lublin area. For 3 months they were assigned to the Sonderkommando, charged with transporting bodies from the gas chambers to the crematorium. From morning to evening, the sole function of this child consisted of exploring the vagina of female bodies in order to search for jewels, gems that might have been hidden there. After 3 months of this occupation, he volunteers for one of the worst mines in Poland. His father and his brothers refuse to leave with him and, a few days after his departure, were gassed as per usual. A year later, saved on a number of occasions by miracle, he presents with quite marked decline, sleeplessness with nightmares, and convulsive seizures. Analysis of his mental and nervous equilibrium by a psychiatrist would have been of great interest.
>
> The other boy, aged 14 years, was arrested at Lyon with his parents. The father was killed in Lyon prison, the mother deported with him was separated on their arrival. He refused to work in any form for the Germans. There would be time, he explains to me, looking at me with a keen eye and a determined look, to learn a job on returning to France, because, at this time, he will be responsible for a young brother, aged 8 years, that the Gestapo had not found.

The Nobel Laureate Elie Wiesel / A-7713, entered Auschwitz aged 15 with his family. He said that he survived because his father was alive: 'And I knew that if I died, he would die'.

[a] Data from US Holocaust Memorial Museum.
[b] Translated by Greg Wilkinson.

The British Journal of Psychiatry (2022) 220, 245. doi: 10.1192/bjp.2021.172

https://doi.org/10.1192/bjp.2022.16 Published online by Cambridge University Press